The Ecology of Uncertainty:

Sources, Indicators, and Strategies for Informational Uncertainty

Christian D. Schunn
University of Pittsburgh

Susan S. Kirschenbaum
Naval Undersea Warfare Center

J. Gregory Trafton
HCI Laboratory
Naval Research Laboratory

Address correspondence to

Christian Schunn
LRDC 821
3939 O'Hara St.
Pittsburgh, PA 15260
USA
email: schunn@pitt.edu

Abstract

Informational uncertainty is a very prominent and multifaceted feature of many complex tasks. Previous work has found some important distinctions for what features of information uncertainty influence behavior, but overall the literature has had little to say about expert performance in a domain on real tasks or has presented views that may be very domain-specific. By carefully examining deep commonalities in three different domains and expert behavior in those domains (meteorological forecasting, fMRI data analysis, and submarine operations), we present general but detailed taxonomies for what are types of sources informational uncertainty, what kinds of indicators experts use to diagnosis the level of uncertainty in a given situation, and what strategies experts use for reducing or dealing with informational uncertainty. The taxonomies structure future research on behavior under uncertainty, education in domains of high uncertainty, and the design of artifacts that support problem solving in domains of high uncertainty.

Studies of behavior in the real world have consistently found that uncertainty has a large influence on behavior. For example, there is a whole subdiscipline of naturalistic decision making focused on judgment under uncertainty (Klein, 1989). While much progress has been made, there is still much to be learned about how uncertainty influences behavior.

Uncertainty is not an undifferentiated concept. There is the subjective uncertainty a person feels, and there is the objective uncertainty in the information a person has, which we call information uncertainty. The central focus of this paper is on informational uncertainty and the way it influences behavior.

Our approach in this paper is essentially one of cognitive anthropology (Hutchins, 1995; Suchman, 1987). Over several years, we have been conducting careful observations of many experts and apprentices working in the three very different domains with high levels of informational uncertainty. We have spent hundreds of hours watching meteorological forecasts on land and at sea in several countries (Trafton & Hoffman, in press; Trafton et al., 2000). We have spent hundreds of hours watching submarine experts, instructors, and students work with submarine operations in high and low fidelity simulators on land, and in submarines at sea (Kirschenbaum, 1992, 2001; Kirschenbaum & Arruda, 1994). We have spent many tens of hours watching fMRI researchers analyze their data (e.g., Trickett, Fu, Schunn, & Trafton, 2000), and we have ourselves begun to design fMRI experiments and analyze fMRI data. We have interviewed dozens of experts in each of the domains about the ways in which uncertainty enters into their domain and how they deal with it. We have attended colloquia and conferences in each of the domains. We have worked with developers (e.g., computer scientists, physicists, and mathematicians) of new visualization tools in each of these domains.

Our central question was the following: Is there a general way in which one can characterize informational uncertainty across three very different complex domains with high levels of informational uncertainty? We divided this question into an analysis of 1) the sources of informational uncertainty, 2) the indicators of informational uncertainty that experts use to track the current level of informational uncertainty, and 3) the strategies experts use to reduce informational uncertainty.

To develop this framework, the three authors each began from a particular domain perspective. We worked independently, each taking one of the three domains that was our own area of expertise. We used our knowledge of the domains based on our many observations and interviews to list the sources of informational uncertainty that existed in the domain, the indicators that experts used in that domain, and the strategies that experts used in that domain to reduce informational uncertainty. This process produced a very long list at all three levels in all three domains. In other words, there were many sources of informational strategies, many indicators that experts used for tracking level of uncertainty, and many strategies for reducing uncertainty. Each list contained many examples that were deeply bound to the richness of the particular domain.

Our challenge was to develop a taxonomy for each of the levels (sources, indicators, and strategies) that captured the essence of the different kinds of sources, indicators, and strategies across the three domains. The question was whether each of the domains would have very different kinds of sources, indicators, and strategies, or whether the three domains would in fact be deeply similar, defined as being able to find instances from all three domains to fill all cells in the taxonomy. In fact, we found that the three domains were deeply similar, and this paper presents the resulting taxonomies (one for each level) with examples from all three domains in each cell. Moreover, our taxonomies provided complete coverage of our lists—in no cases have we deleted sources, indicators, or strategies that we generated initially from the domain perspective.

By developing a unifying taxonomy for each level, we have developed a much better understanding of the cognitive ecology of complex domains with high levels of information uncertainty. This improved understanding will allow for more structured comparisons of performance across domains and structured studies of how experts deal with informational uncertainty. More importantly, we found many interrelations between the different taxonomies. Some indicators were particularly tied to some sources. Likewise, some strategies were particularly tied to some indicators and some sources. Examples of those relationships are provided in the taxonomies. We will highlight some examples in the general discussion.

There are several taxonomies of uncertainty types in existence. Some come from psychology and judgment and decision-making research (Berkeley & Humphreys, 1982; Howell & Burnett, 1978; Kahneman & Tversky, 1982; Krivohlavy, 1970; Lipshitz & Strauss, 1997; Musgrave & Gerritz, 1968; Trope, 1978). Others come from a broad array of particular disciplines, such as geography (Abbaspour, Delavar, & Batouli, 2003), ecology (Regan, Colyvan, & Burgman, 2002; Regan, Hope, & Ferson, 2002), finance (Rowe, 1994), law (Walker, 1991, 1998), acoustics (Egan, Schulman, & Greenberg, 1961), medicine (Brashers et al., 2003; Hall, 2002), consumer choice (Sheer & Cline, 1995; Urbany, Dickson, & Wilkie, 1989), driving behavior (Vlek & Hendrickx, 1988), educational research (Webster & Bond, 2002), negotiation (Bottom, 1998), military tactics (Cohen, Freeman, & Thompson, 1998), and statistics. The taxonomies from the disciplines are typically armchair analyses rather than from observations of experts in the field, and are based on analysis of a single discipline. None have worked across disciplines to find the common ways in which experts think of uncertainty, the way they diagnosis the presence of the uncertainty, and the strategies they use to deal with or reduce uncertainty in problem solving.

Three Chosen Domains for Study Informational Uncertainty

We examined uncertainty in the submarine sonar domain (following the thread from passive sonar to command via target motion analysis), the military Meteorology and Oceanography (METOC) forecasting domain, and a scientific domain (functional Magnetic Resonance Imaging, or fMRI), representing a broad range of types of tasks, kinds of training of experts, kinds of physical display setups, and time-pressure of the underlying tasks. By examining multiple rich domains, we are better able to determine the general features of problem solving with uncertainty without sacrificing the complexity that is found in real domains.

fMRI. The goal of Functional Magnetic Resonance Imaging (fMRI) is to discover both the location in the brain and the time course of processing underlying different cognitive processes. Imaging data is collected in research fMRI scanners hooked to computers that display experimental stimuli to their human subjects. Generally, fMRI uses a subtractive logic technique, in which the magnetic activity observed in the brain during one task is subtracted from the magnetic activity observed in the brain during another task, with the assumption that the resulting difference can be attributed to whatever cognitive processes occur in the one task but not the other. Moreover, neuronal activity levels are not directly measured, but rather one measures the changes in magnetic fields associated with oxygen-rich blood relative to oxygen-depleted blood. The main measured change is not the depletion due to neuronal activity but rather the delayed over-response of new oxygen-rich blood moving to active brain areas, and the delay is on the order of 5 seconds, with the delay slightly variable by person and brain area. Data is analyzed visually by superimposing color-coded activity regions over a structural image of the brain (see Figure 1a), looking at graphs of mean activation level by region and/or over time (see Figure 1b) or across conditions (see Figure 1c), or looking at tables of mean activation levels by region across conditions (see Figure 1d). Elaborate, multi-stepped, semi-automated computational procedures are executed to produce these various visualizations, and given the size of the data (gigabytes per subject), many steps can take up to several minutes per subject. Inferential statistical procedures (e.g., t, ANOVA) are applied to confirm trends seen visually. fMRI is considered to be an academic domain with a high degree of uncertainty in the data. Several sources of uncertainty contribute to the over high level of uncertainty. First, one cannot perfectly control which thought processes of the subjects in a given condition, and off-task thoughts will cause other areas to become active. Second, the spatial resolution is somewhat coarser than neurons or even assemblies of neurons, which are ultimately thought to be the object of study. Third, the process of measuring magnetic fields inside a solid 3-D object (the human head) is a difficult process that can be systematically biased by a variety of factors (e.g., the closeness to the skull, deformations in areas near the nasal passages due to breathing). Fourth, the processes of neurons themselves are thought to be stochastic. Fifth, the neuronal processes happen at a faster pace than the temporal resolution of fMRI. Moreover, to deal with the measurement noise, the analysis of the data typically averages data across several seconds of time or across many trials. However, none of this uncertainty is displayed in the images used in data analysis.

Submarine Sonar Operations. Uncertainty is inherent in the submarine world. Initially, the basic sensory data are uncertain due to the nature of the primary sensor, passive sonar. The causes of this uncertainty are the ocean environment. Sound is not transmitted in a straight lines through ocean water. It is bent, reflected, and refracted by changes in temperature, pressure, and salinity. In addition, it interacts with the bottom and the sound made by quiet “targets” can be lost in the noise caused by wave action, ocean creatures, and large ships, etc.), Often, even the identity of the noise source is uncertain, both because of high background noise and because of similarities among various classes of man-made noise sources.

Secondly, uncertainty is inherent due to the under-determined nature of target motion analysis (TMA) algorithms for passive, bearings-only measurements. Passive sonar provides only two parameters, bearing (direction) and bearing rate (rate of change in the bearings), with which to compute three parameters—course, speed, and range to the noise source. There are numerous TMA algorithms. Selecting the correct algorithm for the current situation is another source of uncertainty and requires an experienced decision maker.

Lastly, the unpredictability of human actions and intentions (both unintentional and intentional deception) contributes to the uncertainty in submarine operations. In submarine operations, uncertainty is the primary cause of error and delayed action. These problems have become even more critical with the emphasis on operations in the crowded and noisy littoral regions.

A variety of displays are used, including displays of the relatively raw sonar information (see Figure 2a), tables showing the results of different TMA algorithms (see Figure 2b), and various spatial displays of range, bearing, and location estimates (not available for public display)

Submariners are acutely aware of the problems caused by uncertainty, both in sonar and TMA, along with the difficulties of tracking perhaps hundreds of contacts. Shipboard, the TMA operator is tracking multiple contacts, correlating track from different sensors, and evaluating the performance of different TMA algorithms. The TMA operator’s work is supervised closely by highly experienced officers and the results are fed to decision aids used by the most senior and experienced officers on the watch. One of those is the Approach Officer (AO) who is responsible for “fighting the ship” in a hostile encounter. His job is to make sense of all the contacts, all the solutions, all the options, and to make a decision. One of the most time-critical and uncertain events for a submarine is a “close encounter” with another vessel. In this case the AO has a very short time in which to assess the evidence and take action. As there is not the leisure for the multiple legs usually required to localize a contact, the AO must make these decision under extreme uncertainty.

METOC. The problem in weather forecasting is that there is considerable uncertainty in Meteorology and Oceanography (METOC) data, both in observations and numerical forecast models. Forecasters must make judgments that account for and accommodate that uncertainty, although uncertainty is not typically displayed in any of their tools. Weather forecasters examine observations, summaries of those observations, and predictive forecast models that use those observations as input. While they do explicitly examine actual observations by examining satellite pictures or local wind-speed, the majority of their information comes from tools that summarize or use those observations. They see satellite images and loops uploaded onto the Internet. They view predictions made by complex models with varying underlying assumptions. The models are also based on possibly unreliable or sparsely sampled observations. Weather is also extremely chaotic (the “butterfly” effect), and no current numerical forecasting model is able to accurately depict the complete dynamic structure of the atmosphere. The associated uncertainty, unreliability, data sparsity, and underlying assumptions are not explicitly provided to the forecasters. The forecasters must infer the uncertainty, either from experience and training, or because the values are not stable across time or across different instances of the “same” data (e.g., different weather models). The data are transformed in one of many ways (e.g., direct transformation, modeling, combination, and multiple representations) and displayed to the decision maker as stimuli.

For the past couple of years, we have been studying how US Naval weather forecasters and Royal Australian Navy forecasters make decisions. The following is brief description of the typical process used to make predictions. First, the forecasters briefly update the synoptic mental model of the weather that they seem to all carry at all times. This they do by reviewing current conditions in the region of interest. Next they turn their attention to prediction. This is the most time consuming and interesting stage. It consists of assessing how accurately the several available models are currently predicting. The trickier the weather conditions, the more likely the models are to disagree. There are no explicit indications of uncertainty in any of the forecasting tools or data displays observed on two continents, but the differences in predictions are evidence of their inexactitude. To evaluate the models, the forecasters made many individual comparisons (1) between model predictions for the current conditions and the actual conditions (e.g., satellite loops) and (2) among models. These comparisons appear to help the forecasters to assess the uncertainties in the models. The first type of comparison assesses the accuracy of the model at the current time. The second assesses the amount of uncertainty in prediction by seeing how much the models differ. Agreement among models increases confidence. Disagreement requires adjustment in the forecaster’s predictions. Figure 3 is a snapshot of a forecaster making such a comparison. Comparisons are made on many variables, across time, altitude, and geographic location. Approximately 25% of the total time-on-task was spent engaged in these comparisons. The probabilistic relationships between the events in the world and the stimulus information available to the decision maker are not usually explicitly provided; most weather visualizations do not explicitly display probabilistic information or even links between co-occurring events. Realistically, even the stimuli do not just show up. The forecaster must search for them or select which information to view, from a much larger set of options. Many of these options are not informative. That is, they do not distinguish between possible events or they provide no new or useful information on the state of the world.

After assessing the situation and models, the forecasters validate their predictions with other forecasters or with previously prepared forecasts. Again, this is a comparison process, but with an emphasis on consistency, not contrasts. Lastly, they prepare their reports, forecasts, and briefs for the customer. The customers for Naval forecasters range from pilots who need detailed flight weather information for the next several hours to tactical and strategic planners who need to know how weather will limit decisions being made for missions in the next 72+ hours. Developing such varied weather products today is a time consuming process that requires extensive experience.

Important differences across domains. The three domains differ along at least four core dimensions on the nature of the tasks done by experts: 1) immediacy, 2) control of information access, 3) malleability, and 4) and number of displays. Immediacy refers to the amount of time the user is typically afforded to make decisions based on the data provided by the display. Submarine decisions are the most immediate with decisions often required in seconds or minutes, METOC-based decisions are required in hours or days, and fMRI-based decisions are often made over days or weeks. Control of information access is the degree to which the artifact can be actively engaged in accessing additional information to reduce uncertainty. METOC is the least amenable to accessing additional information because primary data sources (e.g., weather balloons) are scheduled and deployed independently of the display system. A submarine system has more flexibility in that controllers can take actions (e.g., change course) that will result in information to the system, while fMRI allows the user to access additional data at the discretion of the researcher. Malleability of display refers to the degree to which the display can be modified to enhance analysis and reduce uncertainty. Submarining is the least malleable with fixed display parameters. METOC is somewhat more flexible than submarining in that changes in model parameters can be changed yet the system is still quite constrained. The fMRI display is often changed in the course of data analysis. Number of displays simply describes the number of displays to monitor and, implicitly, the amount of information provided in the total array. METOC displays have the most display sources followed by submarining and finally the fMRI display, which is typically a single monitor.

Our three domains are not representative of all domains nor all performance levels. We do believe, however, that our three domains are very representative of experts conducting complex visual data analysis with high levels of informational uncertainty. Moreover, our analysis focuses on informational uncertainty rather than on decision-making uncertainty, and this aspect of our domains is likely to be general. Thus, we believe that our taxonomy does include the general distinctions that can be found in the other taxonomies developed in various other domains. Interestingly, our taxonomies are noticeably more detailed, likely reflecting our focus on domains with high uncertainty. In the general discussion, we will return to a comparison with other uncertainty taxonomies.

Sources of Informational Uncertainty

At the top level of our taxonomy, we divide information uncertainty into four broad classes: Physics uncertainty, Computational uncertainty, Visualization uncertainty, and Cognitive uncertainty. Here we describe and decompose each of these uncertainty classes, providing examples of each from all three domains to show that these sources are generally prevalent sources of information uncertainty. Figure 4 shows a hierarchical graph of the taxonomy.

Physics uncertainty

Sometimes uncertainty in the information available for the decision maker comes from uncertainty in the raw measured information itself. This measurement uncertainty subdivides into three further subtypes, roughly corresponding to absence of signal being measured, noise/bias in the signal, or noise/bias in the way the signal is being transduced.

Not measured uncertainty. In many complex domains, the measured signal is inherently ambiguous because key information is not measured, either because the information is not recorded or the information is not measurable (with existing technology). Unrecorded information can be a general state of the equipment, or a temporary error producing some missing data points. At the simplest level, observations have a finite precision; the "remaining" precision is basically unmeasured uncertainty, which can then propagate throughout the system.

In fMRI, the most central unmeasured information is actually the brain activity itself. Instead, fMRI measures the blood oxygenation levels, which responds with variable lag by person and brain region to brain activity levels. The variable lag is technically the unmeasured element that causes uncertainty. Because it is not always clear which regions will be important and it is not always feasible to measure the whole brain, some (determined later to be important) brain regions can be unmeasured in a given study. Similarly, the researcher must make decisions about how densely to measure an area (i.e., how far apart the measured slices are). Finally, because the measurement hardware/software are different than the experiment hardware/software it is often hard to measure the exact lag between the start of an fMRI data sequence stream and the start of the experiment program (and thus the presentation of various items to the participant).

In submarine sonar, the number one source of uncertainty is that the (passive) sonar signal provides no range (to target) information directly—only angle of signal and movement of the angle over time are measured directly. Without the range information, there are an infinite number of course, range, and speed combinations that could produce the same signal. The unknown identity of the sound contact is another key unmeasured piece of information (e.g., another submarine, a boat, a whale). Two other frequent pieces of relevant but unmeasured pieces of information are the sea bottom profile (which changes the sound signal) and the precise and current sound–velocity profile (SVP), which also alters the signal. The SVP is available in a historical database, analogous to climatological data, and is measured once daily, wherever the boat is at that time.

In METOC, weather models make very detailed predictions using previous measurements. These measurements can be very sparse due to weather satellite blind spots and other measurement device placements (especially at sea), which introduces considerable uncertainty in the predictions due to not measured information. Of course, point observations may not be very representative of what happens outside the point or how far that point is propagated throughout the numerical model.

Signal noise uncertainty. The information signal, before it reaches the measurement device, can have stochastic variability, unknown levels of bias, and extraneous signal sources, which introduces additional information uncertainty.

In fMRI, electrical-magnetic noise in the surrounding environment, the participant's body beyond the brain (especially the eyes), and the way the signal moves from the brain through the skull introduces considerable signal noise uncertainty. The bias produced from some of these effects is only partially predictable (e.g., there are weaker signals from deeper in the brain, and there are artifactual signals near the eyes due to eye-movements).

In submarine sonar, there is considerable signal noise uncertainty due to the sound signal bending as it travels through water layers of temperature and salinity differences and even bounces off the sea bottom, thermal layers, and sea surface. Various biological (e.g., whales, snapping shrimp) and non-biological (e.g., fishing ships) ocean inhabitants introduce other sources of noise into the sound signal.

In METOC, cloud base height (the bottom of the cloud deck) and cloud height are measured by a device called a ceilometer. The ceilometer measures particulates in the atmosphere and thus smoke, dust, and even precipitation can lead to an inaccurate reading. The reading may also be erroneous if there a small break in the clouds above the instrument. Other measurement devices such as the human eye, satellites, pilot reports, and balloons can be misled by one or more of these phenomena that distort the signal. Such problems are not the fault of the measurement device, but inherent in a noisy signal.

Transduction uncertainty. The third subtype of physical uncertainty is a function of the measurement device itself. In the process of transduction (converting incoming physical energies such as light, sound, magnetic fields, and heat into data), a measurement device may reduce the quality of the incoming information to thereby introduce another source of uncertainty. At the simple level, a device may simply misread an incoming signal, either through poor calibration or high added variability. At the complex level, a transduction device may compress dimensionality, dramatically reducing information being measured.

In METOC for example, a radiosonde may transmit uncertain information about relative humidity under some situations. A radiosonde is a weather balloon that is released into the atmosphere, rises approximately 30 km, and transmits information on temperature, humidity, and pressure by radio as it rises. When the radiosonde gets very dry or very cold, its ability to transmit accurate relative humidity degrades by more than 10%. Interestingly, the National Weather Service allows a 5% error rate (signal uncertainty), but certain conditions cause the error rate to rise well beyond 5%.

In submarine sonar, one of the measurement devices (towed array) groups all sound information on a common cone together (e.g., it treats 10° to the left the same as 10° to the right or 10° straight down or 10° straight up). Moreover, the thickness of the region can subtend several degrees in some regions. This transduction process adds considerable uncertainty about the location of the detected target.

In fMRI, considerable care is taken to make sure the measurement device is as accurate as possible. Yet currently, the level of measurement fidelity is not as high as researchers would typically like. Newer measurement devices with stronger magnets and various procedural improvements increase the measurement fidelity, but the measurements are still not as spatially precise as some researchers would like and not as temporally precise as other researchers would like. Another issue is that many fMRI researchers still use medical scanning equipment in hospitals, which are used most of the day for clinical work and for research primarily in the evenings. This high-use equipment can move out of alignment and this possibility introduces uncertainty into the data. Moreover, some fMRI experiments use a combination of methodologies (e.g., eye-tracking or pupilometry or EEG together with fMRI), and the other equipment can introduce calibration problems into the fMRI measurements.

Computational uncertainty

Once large amounts of data are measured, most domains use a number of potentially elaborate computation procedures on that data before the human sees the data. These computational procedures can add new sources of uncertainty in three general ways.

Future prediction uncertainty. Data is collected at a certain point in time, and the world continues to change beyond that point in time. The computational procedures either make no corrections for these changes or they make only partially accurate corrections, and this introduces a potentially large source of uncertainty.

METOC has the largest degree of future prediction uncertainty because the forecasters must make predictions for times hours, days, weeks, or even months away. Over the longer time periods, uncertainty increases even further because prediction-relevant information about topography, vegetation, and land use can change radically through earthquakes, deforestation, and construction.

In submarine sonar, the task is to predict the location of an enemy submarine some (usually relatively short) time in the future. Even if the exact current location, course, and speed are known, the enemy submarine could change course and speed in the future, and this introduces uncertainty into the predictions.

In fMRI, brain structural images are collected typically only at the beginning of a scanning session. These structural images are crucial for interpreting the functional location of a given area of activation. Yet, the head often moves slightly during the experiment and the brain itself undergoes minor deformations over time (e.g., the areas near the nasal passages deform during breathing). These changes over time since last measurement introduce uncertainty into the analyses.

Statistical artifact uncertainty. Many statistical algorithms/procedures are applied to complex domains to deal with physics or future prediction uncertainty. These statistical algorithms have the potential of introducing artifacts in the displayed data. That is, certain features being displayed may be the result of the statistical algorithm used rather than a reflection of reality. We distinguish two main subtypes of statistical artifact uncertainty.

The first subtype is aggregation/smoothing uncertainty, which removes real features from the data. Statistical algorithms typically try to find a simpler, smoother underlying explanation of the data, filtering out "noise" and possibly aggregating data across larger spatial or time scales than when the data was collected. However, sometimes the features being filtered out are in fact real features of the external world rather than noise. Thus, the information in a display may be more feature-sparse than the real world.

In fMRI, the search for activated regions uses a statistical thresholding procedure that can require sustained activation across time and across adjacent voxels (the smallest unit of volume measurement). Thus, single, very small areas of activation are removed by this procedure.

In METOC, microclimates occur in protected valleys, atop high mountains, adjacent to large bodies of water, etc. These may be more sharply defined than the analysis model predicts. For example, the weather forecast for Newport, Rhode Island, which sits on an island at the mouth of Narragansett Bay, uses the forecast for Providence, 25 miles away, but is often insufficiently adjusted for the sharp influence of the ocean at the coast.

In submarine operations, the algorithms do heavy smoothing to adjust for noise in the signal, and this smoothing process means that the current estimate for course, range, and speed can be off when the signal goes through a quick change, for example during a maneuver by own ship or target.

The second subtype is statistical assumption uncertainty. To infer underlying properties or to further filter out the noise, the computational algorithms/procedures often depend upon certain statistical assumptions. These assumptions can be globally very accurate (i.e., correct most of the time), but locally inaccurate (i.e., untrue for a particular time and place). These assumptions can introduce features into the data that are not actually present, as well as remove features that are actually present. The most common statistical assumptions across domains are linearity (changes over time and space are linear) and persistence (things tend to stay the same over small units of time).

In METOC, a very good predictor of future weather is current weather. Thus, persistence is included into weather models and all weather models make heavy use of the current weather in their predictions. Yet on a local basis, some areas can change weather much more quickly that the models would normally predict. As another example, when measurements in a particular area are sparse (e.g., rain data in a low population area), linearity is sometimes assumed to predict rain data in the locations between actual measurements. This assumption can often make it look like it rained in places that in fact received no rain.

In fMRI, there are algorithms that correct for head motion. The algorithms often assume linear movement over time (e.g., a head gradually sinking into a pillow). Yet, head movements can often be non-linear (e.g., sudden or back-and-forth), and the motion correction algorithms can then move locations of activated areas incorrectly or remove areas of actual activation. As a more complex example, some analysis procedures morph the areas of activation of a given participant into a common reference brain (e.g., adjusting for overall differences in brain size or locations of large reference brain structures). This morphing procedure makes many linearity assumptions in how small regions should move and these assumptions can be incorrect for particular areas.

In submarine operations, one very core assumption (computationally and visually) is that the target is keeping constant course and speed over the recent window of time. Clearly, when the target undergoes a course change and that course change is not yet detected, the algorithms' estimates will be systematically incorrect. As a more complex example, there are sound filtering procedures that make statistical assumptions about how what kinds of sounds can be removed. These assumptions can be incorrect in certain environments and thereby introduce incorrect estimates in the algorithms.

Fast+cheap uncertainty. In many complex domains, there exist elaborate algorithms that have very high levels of accuracy. Unfortunately, even on modern high-power computers, these elaborate algorithms require considerable time to complete, and often longer than the problem solver usually wants to wait. Thus, in these complex domains, there exist many algorithms that are more approximate in accuracy but much faster to run, and capable of being used in real time by the problem solver. The use of these fast+cheap algorithms/procedures introduces additional uncertainty into the problem solving. Many times, one can think of the impact of the fast+cheap algorithms as an aggravation of the statistical assumption of uncertainty.

One special case of fast+cheap algorithms is anytime algorithms. Rather than displaying nothing until the full computations are completed, anytime algorithms display the current best guess so that the problem solver can make quick (although more approximate) judgments if necessary. Anytime algorithms, until they complete, introduce the same kind of computational uncertainty as other fast+cheap algorithms.

In the submarine world, there exist algorithms that take into account very accurate maps of the sea bottom and recent measures of sea temperature and salinity profiles to model the bending and bouncing of sound to determine much more accurately the like sources of a given sound target. However, current submarines use much faster and approximate algorithms.

Similarly in METOC, there are weather models that use very exact, micro models of the physics of weather change, including detailed local maps, but it is not possible for most forecasters run such models for even medium-sized regions. In fact, unless one has access to supercomputers, the time to run the algorithms for even medium-sized regions would introduce additional future prediction uncertainty because it will have been a long time since the last data input to the model. Instead, forecasters use a mixture of very approximate global weather models and more exact (but still relatively approximate) meso-scale weather models.

In fMRI, data analysis is very much an iterative process. Before conducting the most detailed and accurate analyses, many initial decisions have to be made (e.g., about whether the experiment should be continued at all, which participants to keep in the analyses, what overall activation thresholds should be used). For these initial analyses, researchers will typically use more approximate but significantly faster analyses.

Visualization uncertainty

After data are measured and processed through procedures and algorithms, the information must be conveyed somehow to the problem solver. The most typical form is through a visualization (e.g., map, table, graph).[1] Sometimes the visualizations introduce informational uncertainty (either through failing to representing relevant information or presenting them in a misleading way). Note that visualization uncertainty is about what information can be logically derived from a visualization not about human errors or confusions about how to interpret information from the visualization. The latter is part of the next class of informational uncertainty, cognitive uncertainty.

Non-represented information. The most obvious form of visualization uncertainty occurs when information (that is logically necessary for developing an accurate understanding of a situation) is missing from the visualization entirely.

In fMRI, variability information is rarely displayed. For example, in a table of activation in selected key regions, there is sometimes no information displayed about variability of activation across subregions, and thus the problem solver should be uncertain about whether the difference is statistically significant. As another kind of example, sometimes key regions of activation may be missing from the current visualization because a subset of the brain slices are being displayed and the key area may be not on the set being displayed.

In METOC, visualizations rarely display variability information, producing a similar kind of informational uncertainty as in fMRI. Some of the measures are very stable and highly sampled, whereas others are not very stable over the time window being displayed and/or poorly sampled, but the visualization provides no information about the certainty of the displayed means.

In the submarine, depth information is absent in the (2-D) displays. The problem solver begins with the assumption that the contact in the same general horizontal plane as own ship, but this assumption is often incorrect.

Composite information uncertainty. Sometimes multiple dimensions that are typically correlated in value or combine functionally are represented with a single-composite measure. When the functional combination and the correlation among values are not perfect, this composite measure introduces uncertainty. The most common composite is one that shows a combination of duration and amount, so that one does not know whether it was actually a very large amount for small percentage of the given time interval or a smaller amount for the full time interval.

In METOC, many of the displays contain the duration/amount uncertainty. For example, the display of rain levels does not tell you whether it was 1cm of rain gradually over 10 minutes, or a onslaught of 1cm in 1 minute followed by 9 minutes of no rain.

In fMRI, one sometimes has tables that show total signal change of an area (relative to baseline), which combines number of activated voxels in the region with amount of signal change in the activated voxels. Thus, one does not know whether just a few voxels are very active or whether many voxels are moderately active.

In submarine operations, there is a number that represent the noise level in the measurements. This measurement is not an instantaneous measure, but rather an average across a moving window. This number is another composite of duration and amount.

Inconsistent information uncertainty. In many complex domains, there are multiple visualizations shown at the same time. Similar to the joke about the man with two watches never knowing what time it is, multiple displays produce uncertainty when the displays do not match.

In submarine sonar target-motion analysis (TMA), the visualization tool shows estimates from different algorithms, and these estimates frequently do not match in the beginning of problem solving.

In METOC, the forecaster frequently examines predictions of different models and these predictions often do not agree in all their details. The task of the forecaster is then to determine which ones are more correct and/or how they should be combined into a single prediction.

In fMRI, there are different ways of slicing the data. For example, one can examine which regions activate by time or which regions activate by condition. Sometimes these different visualizations do not agree.

Cognitive uncertainty

Humans are a key part of the information system. They act as encoding devices, information storage/retrieval devices, and procedure enactors. Correspondingly, they are also a common source of information uncertainty, as possible errors can be introduce in encoding, retrieval, or procedural enactment. This source of uncertainty can reside in the focal problem solver, or it can come from the information that team members provide.

Note that cognitive uncertainty refers to uncertainty due to possible errors in the cognitive processes of the problem solver that produces information for decision making. It does not refer to whether the problem solver believes there is uncertainty in general or in a particular instance. In other words, it is still a kind of information uncertainty.

That human cognitive processes are fallible is not news. In this section, we decompose the various ways in which failings of human cognitive processes introduce additional informational uncertainty, along with a brief summary of what factors have been shown to be particularly important determinants of performance. We provide examples from observations of expert problem solvers to highlight that cognitive uncertainty is a very important source of informational uncertainty even in expert problem solvers.

Perceptual error. Information from the measurement devices and computers is conveyed to the problem solver, typically visually. The problem solver must then perceive this information, which includes a transduction process (e.g., from light or sound to the retina or inner ear), an attention process (so that complex objects may be encoded), and a pattern recognition process (to decide to what category the observed patterns belongs). Each of these steps of perception can introduce errors. In terms of transductions, the human ear and eyes have minimum detection levels as a function of sound and light frequency and intensity levels (i.e., some sounds/lights are not perceivable). There are also limits in the ability to perceive differences between intensities, typically somewhere in the 5 to 10% change range (known as Weber's law). The attention process acts as a filter—only a limited amount of information can be perceived at once (Treisman, 1969; Wolfe, 1994; Wolfe, Alvarez, & Horowitz, 2000). If attention does not move to an element before it is disappears, errors of omission occur. The pattern recognition process is highly influenced by experience (Biederman, 1987; Polk & Farah, 1995) and expectations (W. F. Brewer & Treyens, 1981), and can produce errors of omission (i.e., failure to categorize a present complex object) and errors of commission (i.e., miscategorizations).

In fMRI, sometimes the researcher will not notice certain areas of activation on the display in front of them, especially if the activation is in an unexpected area. Sometimes researchers misperceive what condition is being displayed in an activation map—the font size on the condition labels is often too small.

In METOC, many of the weather models show quantitative data by displaying a color legend (see Figure 5). Unfortunately, different color schemes can cause different perceptual illusions or cause some colors to be perceived incorrectly. For example, simultaneous contrast can occur if a color is surrounded by particularly dark or light colors, making it difficult to determine the actual color or compare that color to the legend in a straightforward manner (C. A. Brewer, 1996, 1997; Ware, 1988). These perceptual illusions can cause perceptual errors, leading to cognitive uncertainty.

In submarine operations, certain perceptual cues indicate a close contact, but these cues can easily be missed in the clutter of a high-contact region such as a shipping lane or the Mediterranean Sea.

Memory encoding error. In many complex problem solving domains, truly impressive amounts of information is often directly displayed in front of a person across large and multiple monitors. However, at any one time, only a very small amount of information is directly perceived. This deep limit on visual input (or other perception input) implies that complex inputs must be stored and retrieved from memory. In other situations, a person examines different displays sequentially, and must integrate information across displays making even more use of human memory. Unfortunately, the human memory storage process (encoding) is far from perfect. Thus, when a person is reasoning about some situation, they must take into account the fallibility of their own memory encoding process.

The human memory encoding process can be error-prone in two different senses: stochastic selection and biased selection. First, it is possible that only some small, fairly unpredictable subset of possible information that a person encounters is encoded correctly in memory. Second, it is possible for only particular aspects of the encountered situation is correctly entered into memory. Highly familiar information that can be grouped into familiar, meaningful chunks, for example, is more likely to be encoded (Chase & Simon, 1973). Information that violates expectations or is perceptually distinctive is more likely to be encoded (Bower, Black, & Turner, 1979). Information of emotional relevance is also more likely to be encoded (Cahill & McGaugh, 1995). Human memory encoding errors include both failures to encode information and misencodings of information.

In fMRI, patterns that occur in regions of the brain that a researcher is more familiar with are more likely to be encoded, whether they be expected patterns (and thus decomposable into familiar chunks) or unexpected patterns (and thus violations of expectations).

In the submarine world, some kinds of contacts are more salient than others. Thus, low priority contacts such as fishing boats might not be encoded. When the Officer of the Deck makes maneuver decisions, he must consider the location of all contacts, both surface and subsurface. Fishing boats are particularly troublesome because they can be trawling nets well behind and below the boat. Thus, failure to encode either the fishing boat itself or the trawling net can cause serious problems.

In METOC, one of the primary goals of a forecaster is to determine which of several different weather models is the one most likely to make a correct prediction of the future weather. In order to accomplish this goal, the forecaster must compare and contrast not only different weather models with each other, but must examine different weather models with "truth" (typically in the form of a satellite image) sometime in the recent past. Most of these comparisons occur when the forecaster can only see one visualization at a time (i.e., it is difficult to geo- and time-reference different visualizations), so the forecaster must, in effect, remember different features of the visualizations and compare those features across weather models and satellite images. Several analyses of forecasters have suggested that this comparison process is a fundamental aspect to the forecaster's ability to make predictions (Kirschenbaum, Trafton, & Kramer, in press; Trafton & Hoffman, in press; Trafton et al., 2000) and interpret the data (Trafton, Trickett, & Mintz, in press). As a direct consequence of not being able to see multiple models at once with satellite images, forecasters forget pertinent features of one visualization or have interference from similar visualizations.

Information overload. A person can consider or be aware of only so much information at once, and many complex problem solving domains involve large amounts of potentially relevant information. Although there are not thought to be hard limits on how much information can be kept in working memory, the more information that is kept in working memory, the harder it is to rehearse all the elements to keep them there (ref). By organizing information into templates for a particular situation e.g., like the method of loci), experts are able to include keep truly impressive amounts of information in working memory. However, even expert working memory has clear limits that complex domains often exceed. The consequence of information overload is that information is temporarily lost.

In fMRI, there are many regions of the brain that can become active, and most thought processes activate a number of areas in the brain. A given participant's brain activation is displayed across 10-30 brain slice images. A given study often has 10 or more participants. Thus, there is a lot of information to keep track of, and experts will often go back over the data to remind themselves of information they had just recently encoded.

In submarine operations, there can be many contacts as there are many places to look (hundreds of displays for bearing, depression-elevation angle, and frequency), the sonarman can easily become overwhelmed. Add the ability to hold the same contact on multiple sensors, and the potential for information overload is obvious.

In METOC, the forecaster has access to thousands of weather visualizations showing different weather models, blends, satellite images, time courses, etc. Different weather models and tools organize information differently. For example some weather models and tools provide information on a specific set of "standard" variables organized by area of interest and time, while other models and tools force the user to select a specific area and time, but allow multiple different variables to be displayed. Because many of these variables interact with other variables at different geographical or atmospheric levels, the forecaster needs to keep track not only of different weather models, but different variables within each weather model. Not surprisingly, forecasters will sometimes get confused and need to refresh their memory of what they saw. Additionally, because of the vast amount of information available, forecasters just cannot examine all the information that may be relevant to a particular forecast.

Retrieval error. Once information is actually encoded in memory, there is no guarantee the information will be retrieved at a later point in time, due to either inference and/or decay processes (Baddeley & Scott, 1971; Gillund & Shiffrin, 1984). Moreover, not only can information fail to be retrieved at all, but also erroneous information could be retrieved instead. In either case, the retrieval errors might reflect a stochastic factor reflecting neural firing variability, or a bias towards remembering particular information (e.g., consistent with current expectations, more recently encountered information, or more frequently encountered information). As a result, a problem solver is then left with additional uncertainty about whether the information being retrieved is correct or whether information not being retrieved might be relevant.

In fMRI, data from participants are often examined one participant at a time because each participant can generate gigabytes of data and the computers (even the high powered clusters used in fMRI data analysis) can often process only one participant at a time. In theory, the processing could be done as a batch command for many participants, but often adjustments to the processing must be done by hand and several of the component programs are currently not setup for easy automation. Thus, the researcher typically examines data one participant at a time and often retrieves from memory for comparisons across participant images. These memory retrievals for across participant comparisons are prone to error, and researchers will bring back on the screen old images (or recompute images) to double-check for such errors.

In the submarine, TMA operators concentrate on one contact at a time. High priority targets receive the greatest proportion of their time. Thus, just as a low priority contacts (e.g., fishing boat) might be subject to encoding errors, they (or the region where the nets might be) are also subject to retrieval errors.

In METOC, forecasters will frequently attempt to retrieve a similar situation to the current forecasting problem. For example, a hurricane forecaster may attempt to retrieve a similar case, but because there are very few tools for this kind of case-based reasoning, the forecaster's memory will likely be quite faulty.

Background knowledge error. Uncertainty can also be caused by failing to bring in appropriate background knowledge that changes the interpretation or prediction for a particular situation. These errors contrast with retrieval errors in that retrieval errors are of episodic information (information about a particular state of the world) whereas background knowledge errors are either a lack of knowledge or failed retrievals of semantic information (knowledge about the general state of the world). In general, the same factors that produce episodic memory retrieval errors also produce semantic memory retrieval errors.

In fMRI, even experts will forget what particular regions on a brain map are. There are several different taxonomies for labeling brain areas, and particular researchers specialize in particular brain circuits. But activation that occurs in a new experiment could involve a region not previously examined by the particular researcher, and they could misinterpret what that activated region is telling them about the observed thought processes.

In submarine operations, SVP patterns are influenced by diurnal heating of the upper layer of the ocean, by ocean currents (e.g., the cold Labrador Current or the warm Gulf Stream), and by the freshwater outflow from river systems. These and other environmental effects can cause contacts to appear closer than they are or to disappear suddenly when they cross one of these boundaries. Although submariners have been taught the effects of ocean environments, they can neglect to take them into account when evaluating the appearance and disappearance of a contact.

In METOC, knowledge about how recently a model was run and what input was given to a model is required to interpret the model predictions, and sometimes the expert is missing the relevant knowledge. Background knowledge is also required to adjust weather predictions, and experts sometimes forget (or not have) the relevant background knowledge for a particular situation. For example, one might not know that a given power plant is off on Sundays and this changes local whether patterns. Or one might not remember that the current year is an El Niño year, and thus adjust whether pattern base rates for the region appropriately.

Skill error. Uncertainty can also arise from procedural errors in the data transformation or interpretation processes. For example, the problem solver may fail to do a mental or external transformation step. Here the uncertainty is whether the transformation was done, not the error itself. Skill errors are more likely to occur in steps that are less practiced (Singley & Anderson, 1989; Woodrow & Stott, 1936), less recently practiced (Kyllonen & Alluisi, 1987), or when hurried (Grice & Spiker, 1979), fatigued (Krueger, 1994), or stressed (Alkov, Gaynor, & Borowsky, 1985; Beilcock & Carr, 2001).

In fMRI, the analysis process has many steps, is quite complex, has procedures that can vary depending on the situation, and is continually changing as new procedures are being introduced.

Many researchers immediately want to use the latest innovations in analysis techniques. As a result, they will put up with having to do many steps by hand, and suffer poorly design and poorly documented software. Moreover, each new experiment can introduce new ways of splitting the data. Thus, the probability of failing to do a step correctly or at all, or mislabeling a condition or component of the data is relatively high.

In the submarine, the main decision maker is usually quite expert, and the equipment is highly familiar and well practiced. But other members of the team can be considerably less expert. For example, novice sonar operators make mistakes on occasion in operating the equipment, assigning trackers, passing data, etc. This possibility introduces considerable uncertainty for the main decision maker.

In METOC, coordination of weather models can be complex and errors occur, even with experts, in this coordination process. For example, predictions from the global weather model are provided as inputs to the meso-scale model. If there are significant errors in the global weather model predictions for a particular region/time, then these errors will have a large impact on the meso-scale predictions. Sometimes forecasters, in their haste to develop a prediction, fail to verify accuracy of the global inputs before running the meso-scale model, or they (mistakenly) assume that their partner already validated the global inputs.

Tradeoffs among sources of uncertainty

While the various sources of uncertainty are logically distinct, there are many functional dependencies in the form of tradeoffs in which reductions of one source of uncertainty come at the cost of increases in another source. The existence of these tradeoffs has several important implications. But to understand the implications, we must first describe the tradeoffs.

While the details of the tradeoffs can vary from domain to domain, there are at least three general tradeoffs. The first general tradeoff is between physical and computational uncertainty. Many computations done to address physical uncertainty (e.g., reduce noise, compensate for missing data, estimate unmeasured quantities), but those computations introduce new forms of uncertainty (e.g., possibly incorrect statistical assumptions or smoothing errors). Similarly, to reduce uncertainty due to possible computation errors, one can rely more heavily on raw measurements, but those raw measurements typically come with more noise, missing data, etc.

The second general tradeoff is between visualization and cognitive uncertainty. On the one hand, one can remove information from the visualizations to reduce uncertainty due to possible perceptual/memory overload errors, but that comes at the cost of increasing uncertainty due to missing or overly compressed potentially relevant information. On the other hand, one can add the potentially relevant or more multidimensional information to the visualizations, but this comes at the cost of increasing perceptual/memory overload errors.

The third general tradeoff is a tradeoff between memory and perception within cognitive uncertainty, through adjustments to visualizations. By placing more information in a visualization or show more visualizations concurrently, uncertainty due to memory retrieval errors and/or background errors can be reduced, but at the cost of increasing uncertainty due to perceptual errors and encoding errors that occur in overly busy visual environments.

The existence of these general tradeoffs has two important consequences. First, they imply that any domain with significant physical uncertainty and volume of information should have all four big types of uncertainty (physical, computational, visualization, and cognitive) because the better solutions to tradeoff situations always involve some form of middle ground in the tradeoffs. Thus, we now have a theoretical basis (in addition to the empirical basis from the current study) for why our results should be general to most domains.

The second important consequence of the tradeoffs is that they imply that no magical bullet (computer support tool, visualization tool, or educational intervention) will entirely solve the uncertainty problem, and that many solutions will improve one source of information uncertainty only at the cost of increasing another source of information uncertainty.

Indicators of Information Uncertainty

Although complex domains always have some level of informational uncertainty in them, the level of uncertainty does fluctuate over time. What kinds of indicators do experts use to track the level of informational uncertainty? Our analysis of the three domains suggests there are four general kinds of indicators that experts can use. We cast the four kinds of indicators in terms of a general process of trying to find meaningful patterns in the data being analyzed (see Figure 6).

See no or unusually weak pattern. Not seeing any pattern or highly noisy patterns in data are good indicators of informational uncertainty. The world around us contains innumerable patterns at various levels, and when we don't see patterns in the data, we have good grounds to be suspicious of what we are seeing—perhaps the sensory equipment is broken or disconnected, perhaps an analysis transformation step was forgotten, etc.

There are two important notes about this indicator. First, different domains can have very different kinds of patterns to be seen, and being able to see patterns in a domain is a part of expertise in that domain. That is, what appears as just noise to the novice may be highly ordered and informative to the expert. The second note is that most complex domains have some level of noise, and another component of expertise is knowing what level of noise is acceptable or normal.

In fMRI, the activation profile (change in activation levels of a region over time) can follow a nice smooth curve, or it can be very jagged, indicating high uncertainty. Alternatively, an activation map (topographic map of areas of activation in a condition) can have clear clumps of activity or there can be many, many small points of activity with no clear organization, again indicating high uncertainty (compare Figure 7a and 7b).

In submarine operations, degree of pattern in the incoming data can be formalized in a numerical measure of signal-to-noise ratio, where the signal is the acoustical strength coming from the object being tracked and the noise is literally the background noise (the acoustical strength of all the other sources of sound, like merchants, fish, etc). When the displayed signal-to-noise ratio is low, informational uncertainty is high. However, there are also more perceptual cases of absent or noisy patterns which produce informational uncertainty. For example, the waterfall diagram (showing sound strength coming from different angles changing over time) can either have a clear single signal being tracked over time, or it can be a fuzzy mess, indicating high informational uncertainty (compare Figures 7c and 7d). Similarly, the dot stack diagram (indicating deviation of location from current predictions over time) can follow a smooth curve or line over time, or it can have a wide spread with near random placement from one measure to the next, indicating high uncertainty.

In METOC, there is much more weather data collected in the U.S. than there is in the middle of the Atlantic Ocean or in other countries (Australia, for example). Weather balloons are released from numerous locations all over the U.S., but much less data collected over the ocean. However, because ground features (i.e., mountains) play a large role in the weather, ground-based forecasting can be trickier than ocean based forecasting (hurricanes being a notable exception). Thus, there is a tradeoff between the amount of data available to the forecaster over different locations and the number of other factors going into a forecast (i.e., terrain effects).

See an impossible pattern. A second important indicator of high levels of uncertainty is when an observed pattern clearly violates domain expectations. The most common form is an observation with values that are out of an acceptable range. However, sometimes the values are out of range for the particular type of situation being examined.

In fMRI, it is not good to see data displays with extra-coronal areas of activation (i.e., apparently activated brain areas outside the skull), and it makes one uncertain about all of the data one is seeing. Alternatively, in various plots of brain activation, when brain areas show systematic deactivation either by area or over time (i.e., the brain area appears to become less active than in a resting condition), this observation is taken as an indicator of uncertainty about the information being displayed.[2] Finally, experts have expectations about how high the percent signal change in a region can be (the average % change in activity level of elements in a particular region of the brain relative to the control condition); if the percent change is too high, then it is likely that there was some problem in the analysis process (e.g., thresholds were set too low), and many spurious regions of activation (essentially noise) are being displayed.

In METOC, detectors of impossible patterns are sometimes automated. There is a program that notices anomalies in the data. Also, weather satellites sometimes have out of range tests in their instruments and will flag a probable error state when those ranges are exceeded (e.g., ground temperature readings that are much too high or much too low). The detection of impossible patterns also occurs in the meteorologist. For example, the meteorologist might also note measured or predicted weather values that are clearly too high or too low either in general or for the current situation (e.g., noting predictions of snow over Georgia in the summer).

In submarine operations, the most common example of detecting an impossible pattern is noticing speed estimates that are much too fast. There are two other interesting variants, however. One kind of algorithm allows one to conduct sensitivity analyses: changing one of the dimensions by hand (e.g., speed, range, or course) and examining the impact on the other predicted dimensions. If changing one dimension does not noticeably impact the others (an impossible pattern of sorts), then one should be relatively uncertain about this algorithm's predictions. This impossible pattern occurs when the algorithm has not yet developed a stable solution in which those theoretically interconnected dimensions are actually interconnected.

See a pattern that mismatches known state of world. Not all data are considered equal; some data are considered to be direct measures of truth whereas other data are considered as much more inferential and subject to issues of uncertainty. Thus, an important indicator of uncertainty is when a pattern from one of the less direct measures mismatches the pattern considered a known state of the world—it makes the problem solver doubt all of the patterns being currently displayed in the mismatching display. The difference between mismatching the known state of the world and being an impossible pattern is that the impossible pattern requires no reference to other data collection or manipulation of the world, whereas the mismatch to the known state of the world does require reference to some other data about the current state that is considered truth. One of the strongest forms of known truth comes from a manipulation of the environment—if the problem solver manipulates the environment, they may have very clear expectations for some of the impacts of this manipulation and these expectations are considered a known truth about the state of the world (cf. Baker and Dunbar (2000)on known control trials).

In submarine operations, there are two interesting variations of mismatches to the known state of the world. First, when an algorithm's predicted bearing rate (the rate at which the angle of approach of a measured sound changes over time) does not match the observed bearing rate, then one should be relatively uncertain about the validity of the algorithms other predicted dimensions (e.g., range, speed). In this case, measured bearing rate is treated as a known, and a model's prediction is treated as more inferential. Second, when ownship (own submarine) does a course change, there is a clear prediction for the new bearing (angle at which the sound is coming) and bearing rate (how quickly the bearing changes over time). If the measured bearing and bearing rate does not change as predicted, then there is great uncertainty about the assumptions (e.g., perhaps the contact is on the left rather than the right as previously thought or perhaps the contact made a change in course very recently, which throws off the algorithms). Here, the known is the amount by which the bearing should change (i.e., a manipulation value rather than a measured value).

In fMRI, good researchers typically include validity tests in their designs—conditions for which the impact on brain activation is so well established that it is treated as a known. For example, one might have a simple visual stimulus and expect to see activation in early visual areas of the brain. In well established areas, one could have more elaborate validity tests, such as a certain kind of manipulation producing a certain kind of activation profile in a given brain area. When the researcher does not see these highly expected patterns (considered the manipulated known state of the world), then all of the data is considered uncertain (e.g., it may be that some errors occurred in one or more of the main transformation of the data).

In METOC, there is less opportunity for manipulation, but some sources of data are considered truth. For example, forecasters will sometimes overlay weather model predictions for a current or past time on top of a satellite image for that same time to see whether the weather model predictions match the current 'truth'. If there are significant mismatches, then the weather model's future predictions are considered more uncertain. As another example, forecasters will compare a weather model's predictions the current time at the current location and with what is observable out the window. In this case, weather data coming directly to the eyes is considered the known state of the world.

See a pattern that is inconsistent across data sources. Finally, problem solvers can also use mismatches across multiple (equal status) data sources or (equal status) analysis methods. When a mismatch across data sources or analysis methods is detected, then data from all sources/methods are considered uncertain until further problem solving resolves which source is more likely to be correct.

In METOC, forecasters will compare different data sources (e.g., NOAA vs. Navy sources) or different weather model predictions (e.g., NOGAPS vs. AVN). The forecasters will be more certain when the data from the different data sources match. For example, when hurricane forecasters attempt to predict where a hurricane will make landfall, they examine multiple weather models. Each weather model makes a prediction about the track of the hurricane. Sometimes, there is close agreement between weather models and sometimes there will be an obvious outlier or group that seems less likely. The forecasters can ignore or minimally weight those outliers in a straightforward manner.

Similarly, in submarine operations, there are several different algorithms for determining course, range, and speed. When these algorithms produce similar solutions, then there is much less uncertainty about their accuracy. The solutions from different algorithms are plotted on charts to allow the submariner to evaluate how well they converge.

In fMRI, multiple methods for measuring activation do exist (e.g., single-cell recording, PET, fMRI, MEG), but it is rare to use them on the same person at the same time. Instead, one might have replications within an experiment, and look to find similar results across the within-experiment replications. Alternatively, one might use different analysis procedures (similar to at some level to different weather models or submarine TMA algorithms) and see whether the different analysis procedures produce similar results.

As a final note, it is important to realize that the uncertainty could stem from human sources of information uncertainty. For example, it could be that there is a pattern, but the person failed to see it. Or it could be that the patterns are consistent across data sources, but the person misremembered or misinterpreted one of the data sources.

Taxonomy of Strategies for Dealing with Uncertainty

Having diagnosed the situation as being uncertain, what kind of strategies do experts in a domain use to deal with uncertainty? In domains with high levels of uncertainty, like the ones we are examining, much of problem solving is focused on reducing uncertainty. The strategies are designed to reduce uncertainty, but they are heuristic and can introduce other forms of information uncertainty (e.g., statistical uncertainty or cognitive uncertainty). Thus, it is not surprising that there are at least seven different kinds of strategies that experts across the three domains commonly use for reducing uncertainty. There is no strict chronological ordering or preference of use among these strategies. The order listed below, however, is roughly descriptive of the order in which strategies are likely to be applied (see Figure 8).

We use the term strategies to refer to the kinds of responses experts use to deal with or reduce uncertainty. We prefer the term strategies to heuristics, algorithms, procedures, or methods, because the term strategies highlights the facts that 1) experts typically have multiple behaviors they can and do use to deal with uncertainty in a problem, 2) the choice of behavior is influenced by situation features, and 3) the behaviors are generally adaptive but not necessarily optimal.

Check for likely errors. When results are particularly surprising and especially towards the beginning of a problem-solving episode, the common first step is to look for likely sources of error. The particular sources to look can depend upon the particular tools one is using—not all tools have the exact same set of common errors. However, we list some typical kinds of errors to look for in each of the three domains.

In fMRI, problem-solvers typically look for common skill errors. For example, problem solvers often double-check the splitfile—the way the data is parsed by condition. Conditions can be mislabeled, copied to incorrect directories, or be out of temporal phase with the brain imaging data. Here, the problem solver will re-examine the process by which the data was generated, and perhaps redo to the splitfile from scratch. Because there are so many steps in fMRI data analysis, these steps can change with new techniques being developed or by situational differences, and some component data analysis programs can be poorly documented, there are often many opportunities for skill errors to occur.

METOC forecasters also will look for skill errors. For example, weather portals will traditionally show multiple variables and times for different weather models. The forecaster can click on a specific product (e.g., an AVN weather model that shows wind speed and direction, temperature, and relative humidity 12 hours in the future) to extract specific and general information from it. However, sometimes the automated process that takes the data and puts it on a web site may need to be reset, so only a subset may appear. The forecaster may see, for example, portal that is missing 4 out of 5 products (e.g., products may be missing for 2, 4, 6, and 8 hours, but may be there for 10 hours). This kind of pattern may be indicative of products that are not appropriately up to date. Alternatively, sometimes a satellite picture or weather model may be mislabeled either on the portal or on the title of the visualization. This type of mislabeling can lead to erroneous forecasts if not found.

In submarine operations, because of the complexity of the task and the structure of the training, there are many opportunities for skill errors in submarining. For example, the novice operator may create a seemingly straight dot stack (the visual indicator of a good solution) but have an impossible speed value. Another common skill error is failure to consider that the contact may have maneuvered. Miss-identification of the arrival path for a sonar signal is another common skill error.

Another important class of errors that problem solvers examine in fMRI is statistical artifact. For example, the motion correction procedure is often checked. When participants move their heads by too much or too often, the correction procedure can introduce misleading activations into the results (e.g., the procedure often assumes linear movement, like a head sinking gradually into a pillow, and this can be an incorrect assumption). The problem solver can examine the results with and without motion correction, and also try looking only at data from participants with very low levels of movement. Another commonly examined statistical artifact comes from the brain transformation process. Many data analysis approaches take the data from individual participants and transform the spatial location of data so that the locations of various brain structures match exactly a canonical brain (using a structural image of the participant's brain). This transformation process is approximate and can make regions of activation look like they are in a different brain area than they actually are. The problem solver can look at individual participant activation profiles to see whether the activated areas are actually similar in different participants.

In METOC, if a forecaster wants to use a mesoscale (define mesoscale if not already defined earlier) modeling tool, she should first check the global model that the mesoscale model is based on (the outputs from the global model are fed into the mesoscale model). In practice, however, some forecasters do not systematically check the accuracy of the global model (because of time constraints or because they simply ‘know’ the global model is valid). If the mesoscale model turns out to be very odd, it may cue the forecaster to check the legitimacy of the global model.

Focus on more reliable sources. Some data sources are more reliable than others, and a common response to high levels of uncertainty is to focus on the more reliable sources/kinds of data. It is worth noting, however, that reducing the kinds of information being examined/displayed can introduce other kinds of informational uncertainty.

In submarine operations, the common approach is to focus on one particular algorithm from the several being displayed. Three factors enter into the choice. First, the expert problem solver will eliminate the outlier algorithm. Second, experts prefer algorithm that allow them to more tightly monitor levels of uncertainty. One algorithm gives the operator more control, which allows them to tightly monitor the level of uncertainty. Another (newer) algorithm has explicit displays of uncertainty levels. Both of these algorithms are commonly preferred. Third, experts use explicit domain knowledge (formalized in lessons learned and doctrine documents) that presents guidelines for when to use particular algorithms. For example, the guidelines might recommend a particular algorithm for particular geometries, particular environments, or particular maneuvers.

In fMRI, problem solvers do not typically have the outcomes of different procedures simultaneously (as in submarine operations), because the procedures are so computationally intensive. However, when using new analysis software, the problem solver may in situations of high uncertainty switch back to using the older, potentially more reliable software. The more common approach in fMRI, however, is to focus on more reliable aspects of the data. For example, the problem solver will typically adjust statistical thresholds in cases of high uncertainty, so that only the most reliable results are displayed. Another common approach is to throw out high noise or low signal participants—in fact, it is not uncommon to throw out 10 to 40% of participants because of high noise or low signal problems. Finally, the problem solver may only conduct analyses that do not depend upon high fidelity (e.g., not looking at time course of activation within interactions among conditions).

In METOC, a forecaster will learn from experience that some weather models are particularly good at certain situations. For example, she may use weather model A when there is a low off the coast of Florida and a large Noreaster (a storm blowing from the northeast) coming in December, but rely on weather model B when there is not a Noreaster. Even more simply, a weather forecaster may decide that a particular weather model is not handling winter weather well at all this year, so may discount that weather models input entirely. Because of the relatively tight coupling between the forecast and truth (i.e., it is usually easy to determine if a forecast was correct), a forecaster quickly gains faith in certain models in certain conditions).

Adjust for known deviations from truth. When some sources of data can be considered the true state of the world, then the expert problem solver often uses this information to adjust clear deviations from truth in more inferential, indirect sources of data. This adjustment can be done externally in software or it can be done mentally by the problem solver.

In METOC, there are cases of both external and mental adjustments. Some software allows the forecaster to move features on the display by hand (e.g., move the locations of predicted lows). Another key activity is adjusting past state descriptions that weather models take as input. But, verbal protocols of forecasters also show that forecasters will mentally adjust features in a predicted model that they consider in need of adjustment.

In fMRI, the manipulations are primarily mental because it is considered unethical to manipulate data subjectively by hand (rather than clearly described and objective algorithms). However, in verbal protocols, in formal presentations, and discussion sections of papers, fMRI researchers will talk about adjustments that are likely necessary to the location of an activated region, based on expectations derived from results of many previous studies.

In a submarine, information comes from different sonar sources, ranging from a sphere of hydrophones in the bow to an array of hydrophones trailed some distance behind the boat. In order to have a common point of reference, the bearings from these sources must be adjusted to a common location, generally within the command center. For some submarines, the adjustment may have to done manually.

Average across sources/analyses. When there are multiple sources of data and they are equally suspect, the problem solver can average (weighted or unweighted) across the sources of data to produce what they consider a more certain outcome. This averaging process, like the adjustment process, can be done in software or mentally.

In submarine operations, the problem solver can mentally average predictions across some of the algorithms. Several of the algorithms do a temporal smoothing, which is essentially a temporal averaging across data. Some of the algorithms allow the user to control the amount of averaging that happens (e.g., what window to use to remove data from before a target's maneuver).

In fMRI, a common approach with high uncertainty is to move from individual participant data displays to group level average data displays, which is done in software. However, when experts examine displays of individual participants, the verbal protocols indicate that they will mentally average the individual results to develop a more certain overall picture.

In METOC, experts will frequently average mentally across different weather models. It is important to note that this mental averaging is not a simple numerical average, but rather is mediated by a qualitative mental model of the overall situation that is developed (Trafton, 2003). Part of this mental averaging also does smoothing over anomalous data points. Recently, forecasters have begun using statistical combinations of different models, a practice that is called ensemble forecasting.

Acquire additional data of the same kind. To deal with uncertainty, often patience is required; the problem solver must merely wait for more data of the same kind to resolve the uncertainty. It is important to note that waiting for more data can introduce additional uncertainty in the form of possible retrieval errors (forgetting what happened before) and future prediction uncertainty (increasing temporal lag since last collection of key data, like just after a maneuver in submarine operations).

In submarine operations, data continues to come in, and the problem solver can often simply wait for the algorithms to converge on a solution.

In fMRI, the researcher will often decide to collect data from additional participants, in order to see whether unexpected trends will continue or noisy data will average out more clearly.

In METOC, the forecaster can examine more time slices of recent data, or even reexamine the same data again to refresh his/her memory of the data.

Gather more reliable data. Some sources of data are less reliable than others, but may be used at a particular point in time because of cost/accuracy tradeoffs or may just happen to be the data that is currently available. When examining highly uncertain data that cannot be resolved through the simpler strategies described above, expert problem solvers often chose to gather more reliable data, which they may not have done prior to this point because of resource management issues.

In submarine operations, the propagation path of sound underwater is affected by temperature, pressure, and salinity. It can also bounce off the bottom and thermal layers. Thus, changing depth to where the acoustic environment is more favorable can provide a better arrival path and generate more reliable data.

In METOC, similar to the course change action, the forecaster can request a weather balloon to be launched, which provides highly accurate and more recent information for a particular small region. Some more detailed models produce more accurate predictions, but are much slower to run, and thus are not typically used first. Finally, the forecaster can go to other weather models or data sources on the Internet. For example, the Navy is particularly concerned with weather forecasting over the ocean, while the National Weather Service is more interested in forecasting over the U.S. Indeed, the problems are slightly different for ocean- and land-based weather forecasting: ocean-based forecasting has less data input, but the ocean in general is more homogenous than the Earth. Both organizations have created and supported weather models that do an excellent job of both ocean- and land-based forecasting, but because of the different emphasis, there are times when a forecaster may trust one weather model over another. For example, a weather forecaster on a ship in the middle of the Pacific Ocean may rely more on a Navy weather model, while a cyclist trying to determine what the weather will be for an upcoming race may rely more on a NWS model.

In fMRI, similar to METOC, the researcher can use more detailed/accurate analysis procedures that are slower to run and thus not used first. Also, the researcher can decide to change the experiment to a more conservative structure and then collect more data. The more conservative experiment may have more trials per condition, collect baseline data more often (to deal with motion correction issues), or somehow make the manipulation stronger. Note that conducting a more conservative experiment often requires giving up some features of original experiment, such as moving to fewer conditions.

Bound uncertainty in final solution. Uncertainty does not always get resolved by the end of problem solving. Weather forecasts must be made in a timely fashion. Papers must be submitted to conferences and journals. And complete resolution of uncertainty is not necessary. The final common strategy that problem solvers use is to give explicit bounds on the uncertainty of their final solution. Note that these bounds on uncertainty are developed throughout problem solving, rather than just being developed at the last moment.

In fMRI, observed results are categorized as clear or marginal. Verbal hedge words are applied to results that are more tentative. For example, the researcher might state that the location of a particular activated region may actually be in this or another adjacent brain region.

In METOC, ranges of values are often given. For example, wind predictions are between 4 and 7 knots. Alternatively, a weather forecaster may say that “winds are variable.” The wind direction may be considered variable if, during the 2-minute evaluation period, the wind speed is 6 knots or less. Also, the wind direction shall be considered variable if, during the 2-minute evaluation period, it varies by 60 degrees or more when the average wind speed is greater than 6 knots.

In a submarine, different situations have different tolerances for uncertainty. When coming to periscope depth, uncertainty is tolerable so long as all contacts on the left are drawing to the left and all contacts on the right are drawing right and at some minimum range. When hunting an enemy target, some uncertainty is permitted because the torpedo has its own active sonar and can locate a target with some reasonable initial information.

Longer-term strategies. This list of strategies focuses on strategies that expert problem solvers can use while they are working on the problem. It is worth noting that there are some longer-term strategies that communities of problem solvers in the domains are developing to reduce uncertainty overall in their domains. These strategies include improving the measurement tools (e.g., so that it is possible to detect the differences between smoke and precipitation), better modeling tools (e.g., using faster computers or more accurate algorithms/prediction models), training (e.g., to reduce occurrence of certain skill errors), and changes to visualization tools (e.g., to reduce information overload and encoding errors).

General Discussion

We began with the goal of developing a richer understanding of the sources of uncertainty that experts in complex tasks face, the indicators they use to diagnosis its presence, and the strategies they use to deal with uncertainty once it is detected. We began from the perspective of three complex and substantially different domains, using many hours of interviews and observations of working experts in the domains as well as knowledge of formal procedures in the domains, to build domain-specific lists of sources of, indicators of, and strategies for uncertainty.

The lists were not armchair analyses of the domains, but rather developed from concrete cases that we observed.

The core question of this paper was then whether one could find a deep similarity across the very different domains at the levels of sources, indicators, and strategies. It could have been the case that different domains, with different kinds of tasks going over different time courses, with experts of different kinds of training, with very different kinds of measurement devices, etc would have much if not most of the sources, indicators, and strategies being fundamentally unique to each domain. However, we found there were deep similarities at all three levels (sources, indicators, and strategies), and that the process of finding these similarities produced taxonomies for all three levels that were substantial elaborations of previous uncertainty taxonomies. Moreover, in no case did we find a type of source, indicator, or strategy for uncertainty that occurred in one domain but not in the others.

Even more interesting for a cognitive/behavioral analysis, we saw similarities in the interconnections between sources and strategies and indicators across the three domains. For example, in all three domains, checks for likely errors (strategy) focus on skill errors (source) and statistical artifacts (source). As another connection, in all three domains, adjustment for known deviations (strategy) follows pattern mismatches known state of world (indicator). Since some of the connections may be more tacit or probabilistic than logical, future work should examine the many ways in which the strategies are linked to sources and indicators. The establishment of the taxonomies is the first step in that kind of work.

What does this deep similarity of our three taxonomies across domains mean? It could be that a clever and/or motivated person could have found similarities in any given cross-domain set of sources, indicators, or strategies. In other words, perhaps the similarity is in the minds of the researchers and not a deep feature of the world. Without getting into a debate about essentialism and human categorization, we think it safe to at least make the following remarks. First, our three domains were chosen to be broadly representative of domains in which experts analyze complex visual data (a broad class indeed), and it is likely from a simple sampling perspective that the majority of the same categories that we found in our three domains would be found in other domains of that general domain-type as well. Second, it is worth noting that human categorization schemes typically contain exception categories and it is not so easy to find general schemes that work so well across three very different sets of input. Third, examples of sources, indicators, or strategies that are specific to just one domain are perhaps not so important for a science of cognition/behavior in domains of high uncertainty. That is, while we may have unconsciously omitted exception cases or unwittingly misclassified some examples to fit our taxonomies, possible missed cases that are not general in type across domains will not be of interest to researchers studying other domains.

Differences with Other Uncertainty Taxonomies

As noted in the introduction, there are a number of other uncertainty taxonomies to be found in various psychology and disciplinary literatures. In many cases, the taxonomies include the same distinctions that we make, although never as many distinctions. Not surprisingly, taxonomies from the disciplines have emphasized physical and computational sources of uncertainty, whereas taxonomies from psychology have emphasized the cognitive sources of uncertainty. However, our taxonomy, in addition to treating a broader range of categories has more individual cases as well. Rather than documenting all of those correspondences, we focus here on common distinctions that are different or at odds with our own distinctions.

External vs. internal uncertainty. Judgment and decision making research has long divided uncertainty into external (aleatory) vs. internal (epistemic) uncertainty (Hacking, 1975; Howell & Burnett, 1978; Kahneman & Tversky, 1982). External uncertainty refers to uncertainty in outside events (either from variability or randomness or other factors outside the individual's control or ability to predict). Internal uncertainty refers to uncertainty whether a belief is correct. Over the years, researchers have found this distinction intuitive and it has provided important insights into factors that influence behavior.

We find the internal/external labels problematic because internal contains both objective reasons to be uncertain and the subjective feeling of uncertainty. On the objective side, our beliefs are often inaccurate, and thus a decision maker has objective reasons to be uncertain of a belief in the same way that they are uncertain about the outcome of a coin toss. On the subjective side, the term internal is often used to refer to our subjective sense of uncertainty. In fact, objective and subjective uncertainty are very much orthogonal. One can be subjectively uncertain about information that is actually objectively certain. One can be subjectively certain about information that is actually objectively uncertain. To make matters worse, one can be subjectively certain about how objectively uncertain a given piece of information is (e.g., 2 miles plus or minus 10%).

Instead, we present a taxonomy that focuses on the objective. Our taxonomy does include various factors that might be part of internal uncertainty, such as perceptual error, memory encoding, retrieval errors, and background knowledge errors. The remaining elements of our taxonomy might be viewed as external sources of uncertainty, although with much more fine-grained distinctions.

Model error versus variability. Some disciplines make a distinction between model errors and stochastic variability (Abbaspour et al., 2003; Regan, Hope et al., 2002; Suter, 1993). Model errors are incorrect statements about a fixed world, and stochastic variability is uncertainty about some quantity that continues to change over time. We also include notions of stochastic variability in our physics uncertainty, and much of our taxonomy is about factors that produce model errors. However, we find the fundamental distinction between model errors and variability to be misguided because of levels of analysis issues. Take the case of weather. At an extremely local level, there is a large amount of stochastic variability. However, at the level at which predictions are often made (temperature over a county), the weather is much more stable. Yet individual observations that are used to produce the regional model are themselves stochastically variable and can on occasion produce a misleading overall picture. Thus, stochastic variability is one cause of model errors, and it does not make sense to treat them as exclusive fundamental categories.

Behavior consequence uncertainty. Some researchers have included some notion of uncertainty about the consequence of a behavioral act by the decision maker (Berkeley & Humphreys, 1982; Miller & Shamsie, 1999; Pich, Loch, & De Meyer, 2002). In fact, to be clear about our definition of information uncertainty, we specifically focus on uncertainty in information about the state of the world independent of future causal actions by the problem solver. Uncertainty in decision-making settings typically adds behavioral consequence uncertainty(what will happen to the world if I do X) to information uncertainty (what is the state of the world or the future state of the world if I do nothing). Our three domains involved tasks that were fundamentally about analyzing complex data rather than making changes in the world, and thus it is not surprising that we did not find behavioral consequence uncertainty in our domains.

Linguistic uncertainty. In group settings, problem solvers exchange information, and the linguistic process is often much less precise that direct read-off from, and this imprecision can introduce another layer of uncertainty (Regan, Colyvan et al., 2002). In our domains, we find that problem solvers typically develop conventions for avoiding linguistic imprecision or present raw perceptual information directly to other problem solvers (e.g., in the form of printouts, or directing attention to computer screens), thus we did not find that linguistic uncertainty played a large role in our domains. However, other domains may well include more noticeable role for linguistic uncertainty.

Uses of our taxonomies

All three of taxonomies will be of use in educational settings in domains with high uncertainty. Formal decompositions are useful for conveying the structure of a domain. In complex domains, situation categories, situation cues, and problem solving heuristics are often not made explicit to novices, which delays their acquisition.

Another use of our taxonomy will be in the design and redesign of visualization tools. Most visualizations contain little explicit mention of information uncertainty. In all three of our domains, there are various ways in which uncertainty information could be included in visual displays. It is only with a detailed understanding of the sources of uncertainty in a domain that one can begin to understand whether such explicit displays will be helpful. For example, it is unlikely that any visualization of uncertainty will capture more than a small number (typically physics or computational sources) of the many sources. Our taxonomy also draws attention to other kinds of sources of uncertainty that may need very different treatment. For example, the skill errors might require the use of a 'verified' flag that is set when typical sources of skill errors have been examined.

A third use of our taxonomies, especially the indicators and strategies taxonomies, is the development of additional automation to support problem solving in domains with high uncertainty. Consider the case of indicators. Each of the four categories of indicators in our taxonomy can be automated to some degree, and thus one now has four different kinds of indicators that one can try to automate in any domain of high uncertainty. For example, one can develop measures of noise as is found already in the submarine domain. One can make use of information theoretic measures of the degree of patterns in data. One can develop measures of mismatch across data sources. One can develop measures of mismatch from a particular source that is considered the known current state of the world. Finally, one can develop measures of deviations from theoretically possible states.

Similarly, one can develop automation for the strategies for dealing with informational uncertainty, and our taxonomy provides a set of general strategies that one might wish to automate. The value of automating the strategies is not as clear as in the case of indicators, because doing automation (like automatic error correction in typing) is always more controversial that informing automation (like automatic underlining of errors in typing). Whether a given strategy should be automated in a given domain will depend upon two factors: 1) how simple the strategy is to automate relative to how effective people are already at implementing the strategy on their own (e.g., people are good at spatial transformations), and 2) how accurate the automated transformations are.

The fourth use of our taxonomy is for research on behavior (cognition, development, organization, etc) in domains of high uncertainty. Many researchers are interested in behavior in domains of high uncertainty, but tend to focus on particular sources, particular indicators, or particular strategies. There is certainly nothing wrong with focusing on particular elements in a complex situation. But our taxonomies help specify the contrast set, the set of alternative elements that might be considered as also contributing to performance and perhaps possible confounds in the research on the given element. For researchers beginning work in a previously unstudied domain with high uncertainty, our taxonomies provide a starting place for understanding the ecology of the presumably important uncertainty aspect of the domain.

Too many categories?

Science generally prefers parsimonious theories, and one could argue that our elaborate taxonomy for sources of uncertainty is not parsimonious. In other words, for understanding cognition and behavior, perhaps it is not necessary to make so many distinctions. The use of indicators and strategies are the behaviors and cognitions that we must explain. One factor, then, in justifying our sources taxonomy is the extent that the various sources are differentially linked to indicators and strategies. We have documented some of these connections, but additional work is necessary here.

Another factor is the extent to which our taxonomy matches the decomposition of sources of uncertainty that the experts make in thinking about their domain and interacting with other experts in their domain. The majority of our distinctions were made by the experts in interviews with them and in our observations of their work practice. If you ask the experts to talk about sources of uncertainty, they would not typically talk about cognitive sources, and indeed, this is why cognitive sources of uncertainty are typically absent from domain-centric taxonomies. However, their language would reflect intuitive understanding of cognitive sources in general and our distinctions in particular. For example, "can't remember", "didn't see", "lost track of", and "made a mistake" are commonly heard distinctions in their problem solving language.

A third factor is the extent to which the categories are useful for educators or designers of artifacts in the domain. As noted earlier, our taxonomies are likely to be generally of use to educators and designers. What about the detailed breakdowns in sources? One could argue that formal decompositions are easier to understand when they have concrete instantiations. Our broad sources taxonomy breaks down into very concrete concepts, and thus should be relatively easy to understand (in contrast to more abstract divisions like epistemic vs. aleatory).

Conclusion

Through an approach of cognitive anthropology, but applied to three very different domains rather than just one, we have developed three detailed taxonomies for what makes information uncertain, how experts can tell that uncertainty levels are high, and what strategies experts use to reduce or deal with detected high levels of uncertainty. We have cast the taxonomy in domain-general terms, but have provided detailed examples of each category from each of the three domains we examined. We believe our work has substantially improved our understanding (reduced our uncertainty) about the complex influence of uncertainty on expert performance in real world tasks.

Acknowledgments

We would like to thank Tony Harrison, LT Tomas O'Malley, and David Jones for helpful comments on earlier versions of this manuscript. Work on this project was supported by ONR grants N000140310061 to CDS, Project B424104 to SSK, and M12439 and N0001403WR200078 to JGT, and by Naval Undersea Warfare Center’s In-House Laboratory Independent Research Program, as Project C623224. The ILIR program is funded by ONR. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Navy.

Footnotes

[1] In some domains, sonification or even proprioceptive-means are used to convey information to the problem solver. We use the term visualization because it is by far the most common means of conveying information. However, our taxonomy actually refers to the more general class of uncertainty due to way in which information is transmitted to the problem solver, and our subtypes of visualization uncertainty apply equally well to other perceptual input streams.

[2] Normally, brain regions are not supposed to become less active than during resting state. In some cases, however, there are reasons why there may actually be deactivation in an area relative to a 'resting' control condition. For example, in some resting conditions, participants may become bored and engage in some verbal thinking like wondering about the point of the experiment or building a grocery list. In experimental conditions that make heavy use of non-verbal thinking, like mental rotation, one could observe systematic deactivation of verbal brain areas relative to the control condition.

References

Abbaspour, R. A., Delavar, M. R., & Batouli, R. (2003). The issue of uncertainty propagation in spatial decision making. In *Proceedings of the Scandinavian Research Conference on Geographical Information Science* (pp. 57-65).

Alkov, R. A., Gaynor, J. A., & Borowsky, M. S. (1985). Pilot error as a symptom of inadequate stress coping. *Aviation, Space, and Environmental Medicine, 56*(3), 244-247.

Baddeley, A. D., & Scott, D. (1971). Short term forgetting in the absence of proactive interference. *Quarterly Journal of Experimental Psychology, 23*(3), 275-283.

Baker, L. M., & Dunbar, K. (2000). Experimental design heuristics for scientific discovery: The use of "baseline" and "known standard" controls. *International Journal of Human Computer Studies, 53*(3), 335-349.

Beilcock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General, 130*(4), 701-725.

Berkeley, D., & Humphreys, P. (1982). Structuring decision problems and the "bias heuristic." *Acta Psychologica, 50*(3), 201-252.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115-117.

Bottom, W. P. (1998). Negotiator risk: Sources of uncertainty and the impact of reference points on negotiated agreements. *Organizational Behavior and Human Decision Processes, 76*(2), 89-112.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*(2), 177-220.

Brashers, D. E., Neidig, J. L., Russell, J. A., Cardillo, L. W., Haas, S. M., Dobbs, L., et al. (2003). The Medical, Personal, and Social Causes of Uncertainty in HIV Illness. *Issues in Mental Health Nursing, 24*, 497-522.

Brewer, C. A. (1996). Prediction of Simultaneous Contrast between Map Colors with Hunt's Model of Color Appearance. *Color Research and Application, 21*(3), 221-235.

Brewer, C. A. (1997). Evaluation of a Model for Predicting Simultaneous Contrast on Color Maps. *The Professional Geographer, 49*(3), 280-294.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*(2), 207-230.

Cahill, L., & McGaugh, J. L. (1995). A novel demonstration of enhanced memory associated with emotional arousal. *Consciousness and Cognition: An International Journal, 4*(4), 410-421.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. (pp. 215-281). New York: Academic Press.

Cohen, M. S., Freeman, J. T., & Thompson, B. (1998). Critical thinking skills in tactical decision making: A model and a training strategy. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 155-189). Washington, DC: American Psychological Association.

Egan, J. P., Schulman, A. I., & Greenberg, G. E. (1961). Memory for waveform and time uncertainty in auditory detection. *Journal of the Acoustical Society of America, 33*, 779-781.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1-67.

Grice, G. R., & Spiker, V. A. (1979). Speed-accuracy tradeoff in choice reaction time: Within conditions, between conditions, and between subjects. *Perception and Psychophysics, 26*(2), 118-126.

Hacking. (1975). *The Emergence of Probability*. Cambridge, MA: Cambridge University Press.

Hall, K. H. (2002). Reviewing intuitive decision-making and uncertainty: The implications for medical education. *Medical Education, 36*, 216-224.

Howell, W. C., & Burnett, S. A. (1978). Uncertainty measurement: A cognitive taxonomy. *Organizational Behavior and Human Decision Processes, 22*(1), 45-68.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA, USA: MIT Press.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition, 11*(2), 143-157.

Kirschenbaum, S. S. (1992). Influence of experience on information-gathering strategies. *Journal of Applied Psychology, 77*(3), 343-352.

Kirschenbaum, S. S. (2001). Analyzing submarine decision making: A question of levels. In E. Salas & G. Klein (Eds.), *Linking expertise and naturalistic decision making* (pp. 189-207). Mahwah, NJ: Lawrence Erlbaum Associates.

Kirschenbaum, S. S., & Arruda, J. E. (1994). Effects of graphic and verbal probability information on command decision making. *Human Factors, 36*(3), 406-418.

Kirschenbaum, S. S., Trafton, J. G., & Kramer, E. A. (in press). Comparative Cognitive Task Analysis.

Klein, G. A. (1989). Strategies of decision making. *Military Review, May*, 56-64.

Krivohlavy, J. (1970). Subjective probability in experimental games. *Acta Psychologica, 34*(2-3), 229-240.

Krueger, G. P. (1994). Fatigue, performance, and medical error. In M. S. Bogner (Ed.), *Human error in medicine* (pp. 311-326). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Kyllonen, P. C., & Alluisi, E. A. (1987). Learning and forgetting facts and skills. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 124-153). Oxford, England: John Wiley & Sons.

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes, 69*(2), 149-163.

Miller, D., & Shamsie, J. (1999). Strategic responses to three kinds of uncertainty: Product line simplicity at the Hollywood film studios. *Journal of Management, 25*, 97-116.

Musgrave, B. S., & Gerritz, K. (1968). Effects of form of internal structure on recall and matching with prose passages. *Journal of Verbal Learning and Verbal Behavior, 7*(6), 1088-1094.

Pich, M. T., Loch, C. H., & De Meyer, A. (2002). On uncertainty, ambiguity, and complexity in project management. *Management Science, 48*(8), 1008-1023.

Polk, T. A., & Farah, M. J. (1995). Late experience alters vision. *Nature, 376*(6542), 648-649.

Regan, H. M., Colyvan, M., & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications, 12*(2), 618-628.

Regan, H. M., Hope, B. K., & Ferson, S. (2002). Analysis and portrayal of uncertainty in a food web exposure model. *Human and Ecological Risk Assessment, 8*(7), 1757-1777.

Rowe, W. D. (1994). Understanding uncertainty. *Risk Analysis, 14*, 743-750.

Sheer, V. C., & Cline, R. J. (1995). Testing a model of perceived information adequacy and uncertainty reduction in physician/patient interactions. *Journal of Applied Communication Research, 23*, 44-59.

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Cambridge, MA: Harvard Press.

Suchman, L. A. (1987). *Plans and situated action: The problem of human-machine communication*. New York: Cambridge University Press.

Suter, G. W. (1993). *Ecological Risk Assessment*. Boca Raton, FL: Lewis Publishers.

Trafton, J. G. (2003). Expertise in complex visualization usage. Unpublished manuscript.

Trafton, J. G., & Hoffman, R. R. (in press). Computer-aided Visualization in Meteorology. *IEEE Intelligent Systems*.

Trafton, J. G., Kirschenbaum, S., S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. *International Journal of Human Computer Studies, 53*(5), 827-850.

Trafton, J. G., Trickett, S. B., & Mintz, F. E. (in press). Connecting Internal and External Representations: Spatial Transformations of Scientific Visualizations. In *Foundations of Science*.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review, 76*(3), 282-299.

Trickett, S. B., Fu, W.-T., Schunn, C. D., & Trafton, J. G. (2000). From dipsy-doodles to streaming motions: Changes in representation in the analysis of visual scientific data. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Trope, Y. (1978). Inferences of personal characteristics on the basis of information retrieved from one's memory. *Journal of Personality and Social Psychology, 36*(2), 93-106.

Urbany, J. E., Dickson, P. R., & Wilkie, W. L. (1989). Buyer uncertainty and information search. *Journal of Consumer Research, 16*(2), 208-215.

Vlek, C., & Hendrickx, L. (1988). Statistical risk versus personal control as conceptual bases for evaluating (traffic) safety. In T. Rothengatter & R. de Bruin (Eds.), *Road user behaviour: Theory and research* (pp. 139-151). Assen, Netherlands: Van Gorcum & Co B.V.

Walker, V. R. (1991). The siren songs of science: Toward a taxonomy of scientific uncertainty for decision makers. *Connecticut Law Review*(23), 567.

Walker, V. R. (1998). Risk regulation and the "faces" of uncertainty. *Risk: Health, Safety, & Environment, 27.*

Ware, C. (1988). Color sequences for univariate Maps: Theory, experiments, and principles. *IEEE Computer Graphics & Applications*, 41-49.

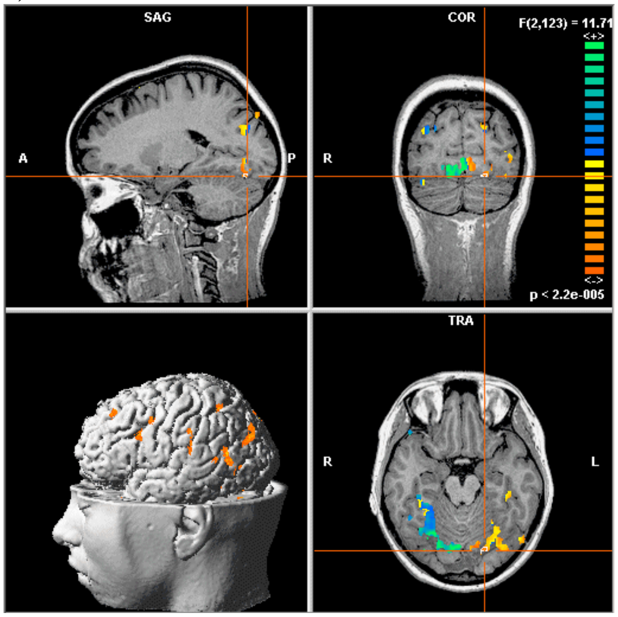
Webster, A., & Bond, T. (2002). Structuring uncertainty: Developing an ethical framework for professional practice in educational psychology. *Educational and Child Psychology, 19*(1), 16-29.

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin and Review, 1*(2), 202-238.

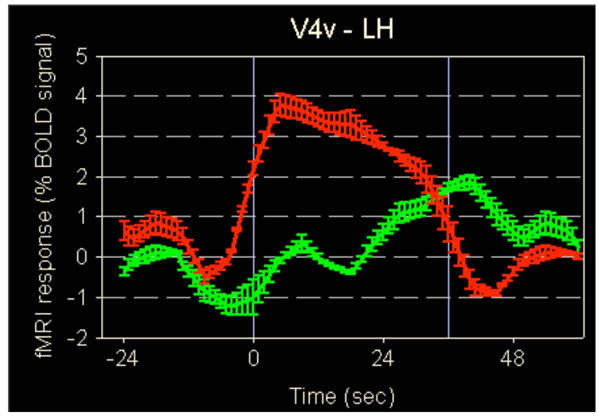
Wolfe, J. M., Alvarez, G. A., & Horowitz, T. S. (2000). Attention is fast but volition is slow. *Nature, 406*(6797), 691.

Woodrow, H., & Stott, L. H. (1936). The effect of practice on positive time-order errors. *Journal of Experimental Psychology, 19*, 694-705.

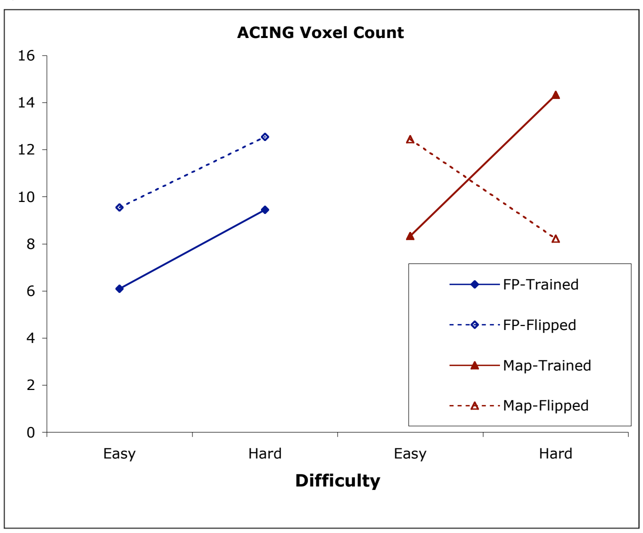
a)



b)



c)

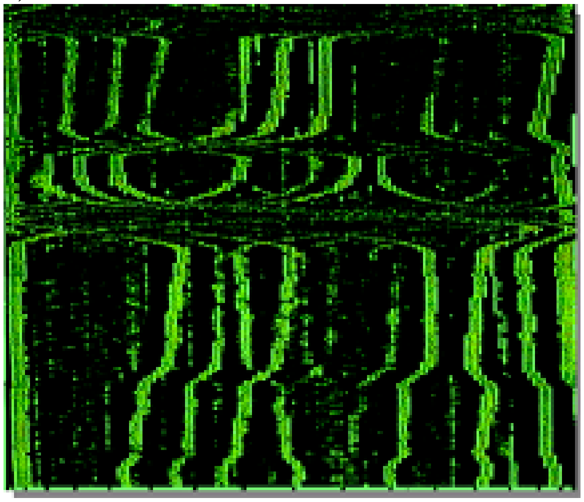


d)

	TrainedEasy	TrainedHar	FlippedEasy	FlippedHard
ROI	Count	Count	Count	Count
1	2	3	4	5
ACING	8.33	14.33	12.44	8.22
CALC	72.33	87.56	75.89	103.22
LCBELL	21.11	33	27.22	28.89
RCBELL	19.44	38.44	28.33	33.67
LCBELL+RCBELL	40.56	71.44	55.56	62.56
LDLPFC	**32.78**	**51**	**38.22**	**40.56**
RDLPFC	**37.56**	**61.67**	**43.44**	**57.33**
DLPFC	**70.33**	**112.67**	**81.67**	**97.89**
LFEF	5.33	8.22	4.67	7.33
RFEF	3.44	5.56	3.78	6.56
LFEF+RFEF	8.78	13.78	8.44	13.89
LHG	12.67	1.44	5.56	1.56
RHG	14	3	6	3.78
LHG+RHG	26.67	4.44	11.56	5.33
LIES	77.56	104.89	84.78	109.44
RIES	59.11	76.56	59.56	81.67
LIES+RIES	136.67	181.44	144.33	191.11
LIPL	26.44	22.22	27.33	20.67
RIPL	19.11	21.89	20.33	20.56
LIPL+RIPL	45.56	44.11	47.67	41.22
LIPS	**54.67**	**76.56**	**61**	**75.22**
RIPS	**37.67**	**64.11**	**48.11**	**72.56**
IPS	**92.33**	**140.67**	**109.11**	**147.78**

Figure 1. Kinds of visualizations examined in analysis of fMRI data: a) degree of activation indicated with a color scale superimposed over a gray-scale structural brain image in three different planar slices and a surface cortex map; b) graph of percent signal change in a brain region as a function of time relative to a stimulus presentation in two different conditions (red and green); c) graph of number of activated voxels in an area as a function of various condition manipulations; and d) table of number of activated voxels in different brain areas (Regions of Interest) as a function of different conditions.

a)



b)

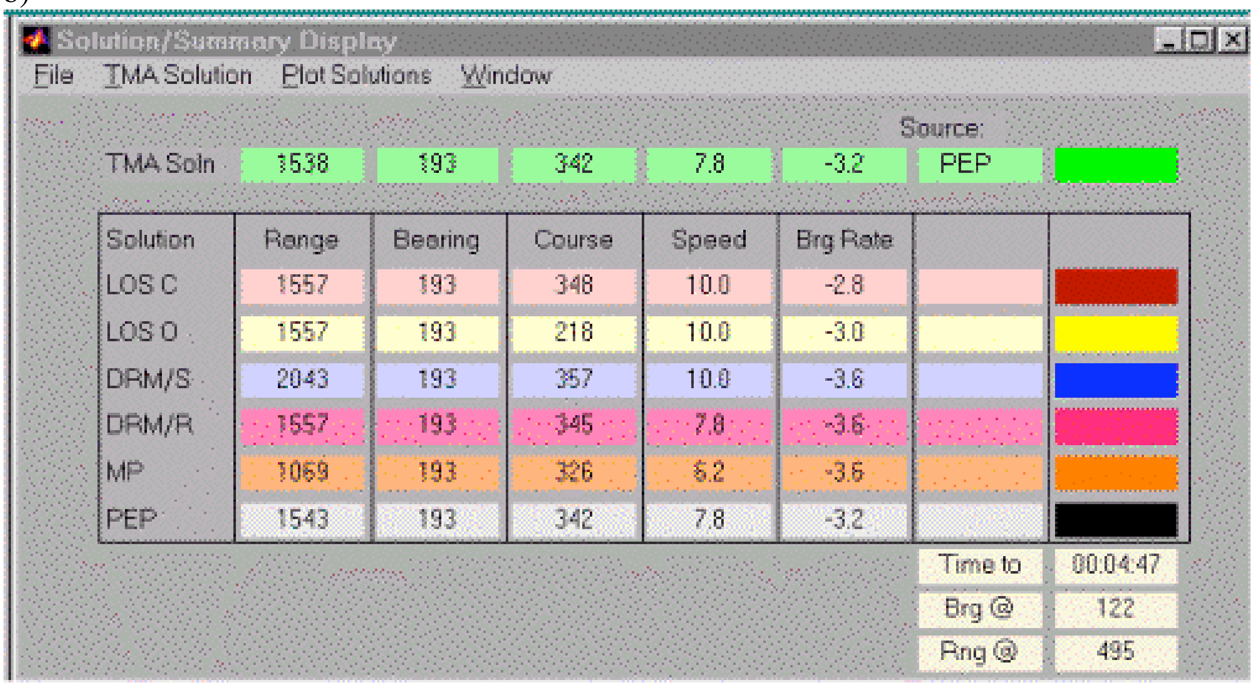


Figure 2. Some visualizations used in passive sonar: a) a waterfall diagram showing the angle of various noise sources across the horizontal axis over time across the vertical axis; and b) a table showing the TMA solutions for 6 different algorithms.

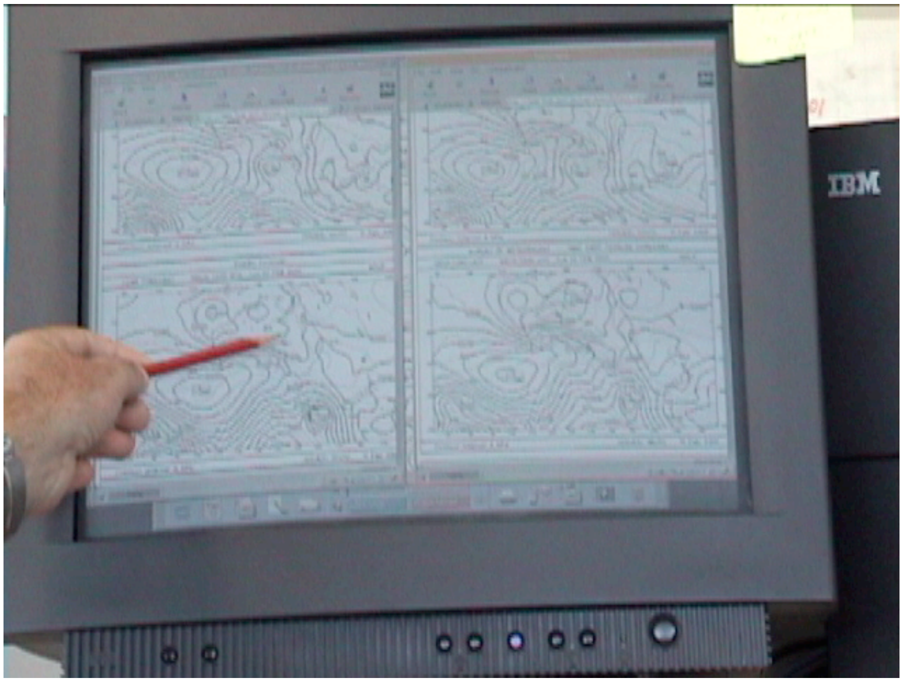


Figure 3. Meteorological forecaster comparing predictions from two models for the same time and location.

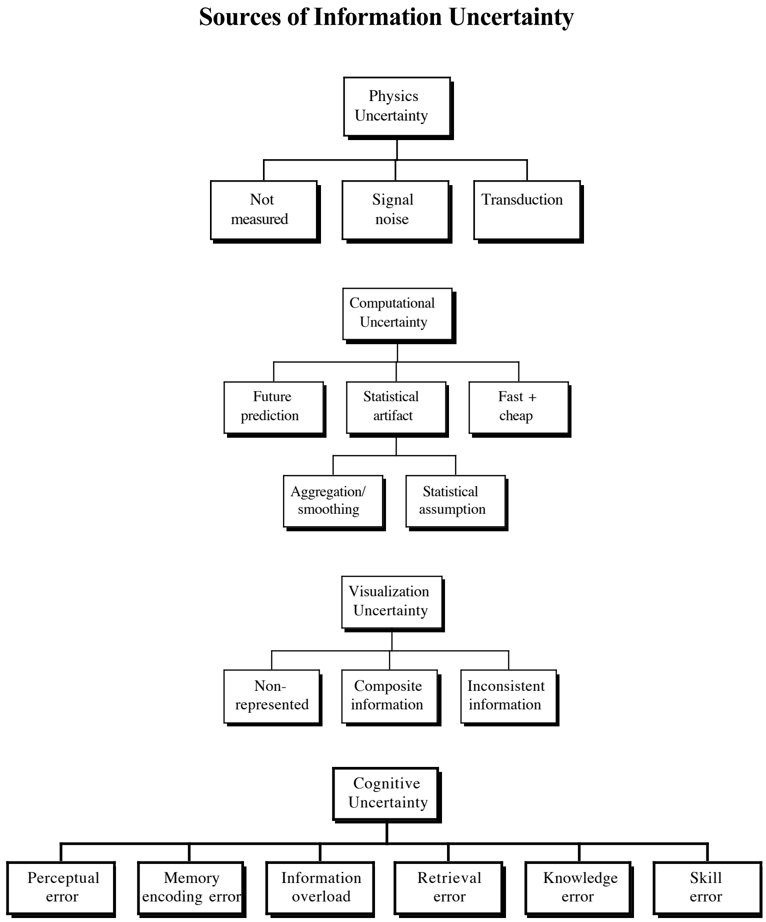


Figure 4. Sources of information uncertainty: Physics, computational, visualization, and cognitive.

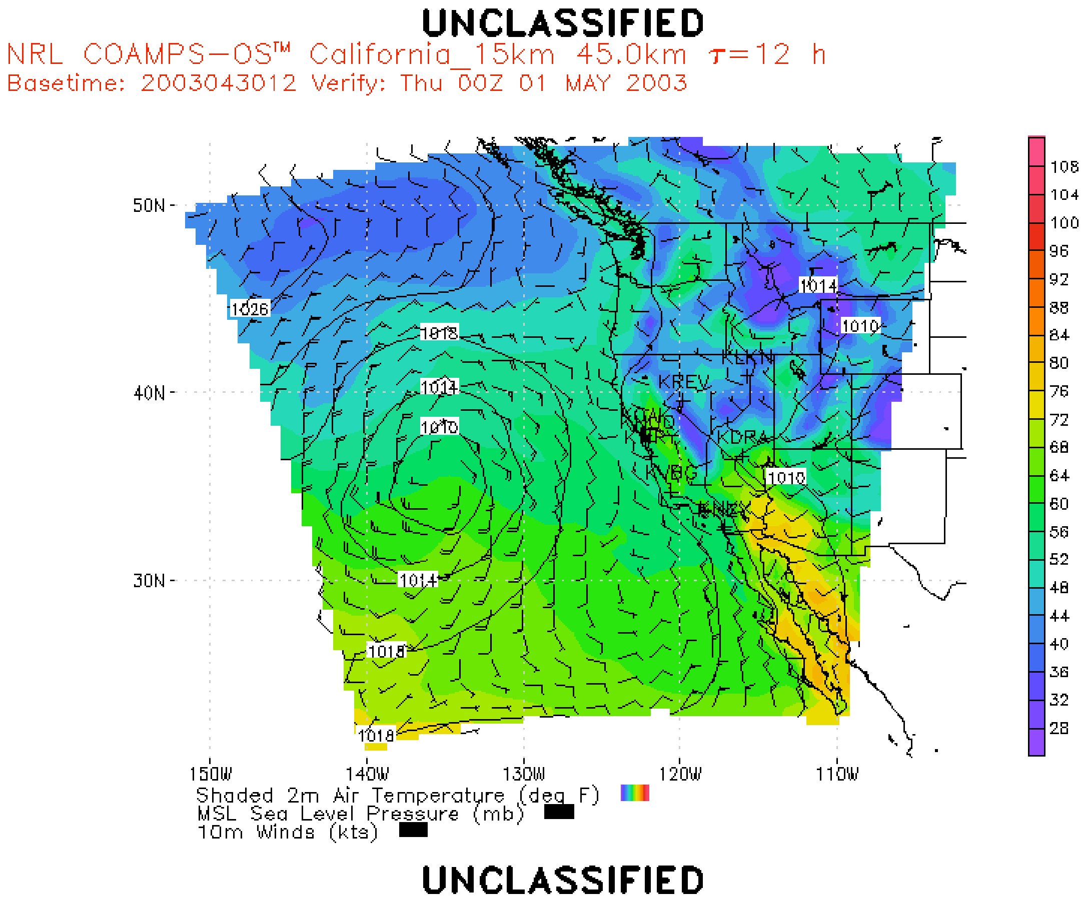


Figure 5: A COAMPS model of air temperature (color coded), sea-level pressure (isolines), and wind speed / direction (wind barbs). The model is a prediction of what these variables will be 12 hours in the future from the time the graphs was generated.

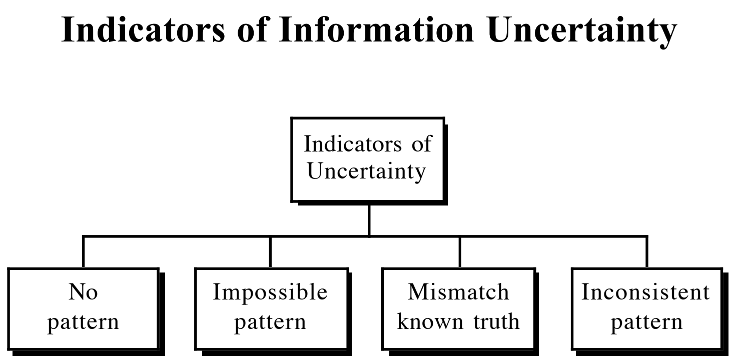


Figure 6. Indicators of information uncertainty.

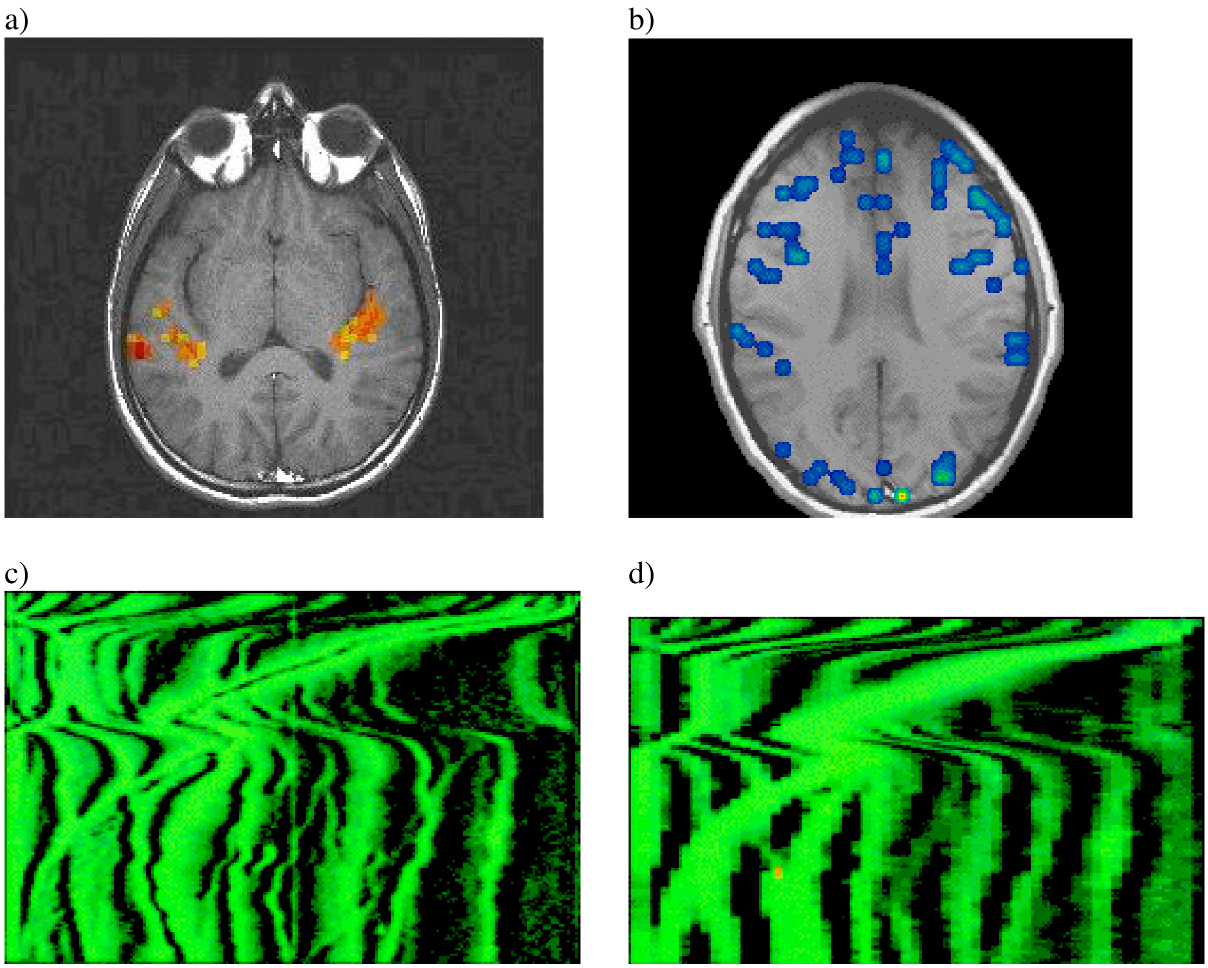


Figure 7. Examples of degree of pattern indicating uncertainty levels: a) fMRI activation map with a strong pattern; b) fMRI activation map with a weaker pattern; c) sonar waterfall diagram with lower noise levels; d) sonar waterfall diagram with higher noise levels.

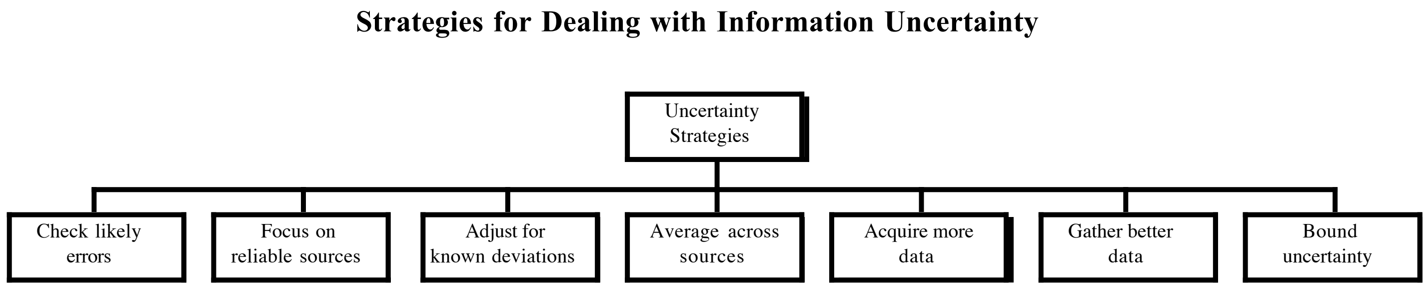


Figure 8. Strategies for dealing with information uncertainty.